

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:15:11 ; Search time 47 seconds  
(without alignments)  
1867.570 Million cell updates/sec

Title: US-09-593-793a-113  
Perfect score: 2861  
Sequence: 1 MVQRLVSVSLRLRRKQAQLL.....AIYFATQVVVDKSLAKYSA 553

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2861	100.0	553	19 AAW71869	Amino acid encoded
2	2861	100.0	553	19 AAW69385	Prostate tumour sp
3	2861	100.0	553	21 AAB28527	Protein encoded by
4	2861	100.0	553	21 AAY82002	Human immunogenic
5	2861	100.0	553	22 AAU69763	Human prostate cDN
6	2861	100.0	553	22 AAU04961	Human prostate tum
7	2861	100.0	553	22 AAU01117	Human prostate-spe
8	2861	100.0	553	22 AAG99002	Human prostate-spe
9	2861	100.0	553	22 AAG62150	Human P501S invent

10	2861	100.0	553	22	AAB74800	Prostate tumour an
11	2861	100.0	553	23	ABG94411	Human prostate tum
12	2861	100.0	553	23	ABG76665	Prostate tumour pr
13	2861	100.0	553	23	ABB77575	Human mast cell re
14	2861	100.0	553	23	ABG61900	Prostate cancer-as
15	2861	100.0	553	23	ABB95222	Human L1-12 protei
16	2861	100.0	553	23	AAU10324	Human PROST 03. H
17	2861	100.0	553	23	AAU82643	Human breast tumou
18	2861	100.0	553	24	ABU71653	Prostate cancer sp
19	2861	100.0	1079	22	ABU74830	Prostate tumour an
20	2861	100.0	1079	24	ABU71860	Prostate specific
21	2601	90.9	710	23	AAU50661	Thioredoxin-ubiqui
22	2596	90.7	595	22	AAU01318	Alpha prepro-P501S
23	1696	59.3	359	24	ABU71887	Human prostate spe
24	1677.5	58.6	530	23	AAU50662	Thioredoxin-ubiqui
25	1417.5	49.5	371	22	AAU69875	Human prostate cDN
26	1417.5	49.5	371	22	AAU01230	P553S splice varia
27	1417.5	49.5	371	23	ABB95335	Human P553S splice
28	1417.5	49.5	371	24	ABU71766	Prostate cancer as
29	1416	49.5	371	22	AAU01362	Human gene 11 enco
30	1416	49.5	371	23	ABG64105	Human albumin fusi
31	1403.5	49.1	400	22	AAU69907	Human prostate pro
32	1403.5	49.1	400	22	AAU01262	Ra12-P501S-E2 cons
33	1403.5	49.1	400	23	ABB95367	Ra12-P501S-E2 cons
34	1403.5	49.1	400	24	ABU71798	Prostate cancer as
35	1287	45.0	255	20	AAW85068	Protein encoded by
36	1287	45.0	255	21	AAB29268	Human prostate-rel
37	1287	45.0	255	22	AAU04205	Prostate-specific
38	1287	45.0	255	23	AAU019084	Human prostate-spe
39	1150	40.2	231	21	AAY54369	Amino acid sequenc
40	1120	39.1	252	21	AAU01423	Human secreted pro
41	1014.5	35.5	326	23	ABB77571	Human mast cell re
42	519.5	18.2	599	22	ABB60709	Drosophila melanog
43	457.5	16.0	748	22	AAU40227	Human polypeptide
44	426	14.9	123	22	AAU69873	Human prostate cDN
45	426	14.9	123	22	AAU01228	P553S splice varia

## ALIGNMENTS

RESULT 1  
AAW71869  
ID AAW71869 standard; Protein; 553 AA.  
XX  
AC AAW71869;  
XX  
XX  
DT 06-JAN-1999 (first entry)  
DE Amino acid encoded by prostate tumour clone L1-12.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone.  
XX  
OS Homo sapiens.  
XX  
PN WO9837093-A2.  
XX  
PD 27-AUG-1998.  
XX  
PF 25-FEB-1998; 98WO-US03492.  
XX  
PR 09-FEB-1998; 98US-0020956.  
PR 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
XX  
(CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
XX  
DR WPI; 1998-609886/51.  
DR N-PSDB; AAV61201.  
XX  
PT Polypeptides comprising immunogenic portions of prostate proteins -

PT used in a vaccine for the treatment of prostate cancer

XX Example 1; Page 82-84; 130pp; English.

CC The present sequence is an immunogenic portion of a prostate tumour protein. The immunogen, or the DNA encoding it, can be used as a vaccine for the treatment of prostate cancer. The immunogen was isolated from a prostate tumour cDNA library obtained by subtracting a prostate tumour cDNA expression library with a normal tissue cDNA library.

XX SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLRRHKAQLLLVNLITFGLEVCLAAGITYVPPLLEVGVEEKFMTVLGIG 60  
DB 1 MVQRLWVSRLRRHKAQLLLVNLITFGLEVCLAAGITYVPPLLEVGVEEKFMTVLGIG 60  
QY 61 PVLGLVCVPLIGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
DB 61 PVLGLVCVPLIGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPHCRQAYSVYAFMISLGGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPHCRQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240  
DB 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240  
QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTLRLLFVAELCSWMLMTFTLYTDFVGEGL 300  
DB 241 CCPCARLAFRNLGALLPRHLQCCMRPTLRLLFVAELCSWMLMTFTLYTDFVGEGL 300  
QY 301 YQGVPRABPTGTEARRHYDEGVRMGSLGLFQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
DB 301 YQGVPRABPTGTEARRHYDEGVRMGSLGLFQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGSGLLPPPPALCGASACDVSVRVVVGSPTEA 480  
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGSGLLPPPPALCGASACDVSVRVVVGSPTEA 480  
QY 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
DB 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540

RESULT 2

ID AAW69385  
XX AAW69385 standard; Protein; 553 AA.

AC AAW69385;

XX 25-MAR-2003 (updated)

DT 08-DEC-1998 (first entry)

XX Prostate tumour specific gene clone L1-12 protein.

XX Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy.

XX Homo sapiens.

XX WO9837418-A2.

XX 27-AUG-1998.

XX 25-FEB-1998; 98WO-US03690.

XX 25-FEB-1997; 97US-0806596.

PR 01-AUG-1997; 97US-0904809.

PR 09-FEB-1998; 98US-0020747.

XX (CORI-) CORIXA CORP.

XX Dillon DC, Xu J;

XX WPI; 1998-480805/41.

DR N-PSDB; AAV58586.

XX Novel human prostate specific tumour protein and fragments - useful

PT for detecting and treating prostate cancers

XX Example 1; Page 87-89; 141pp; English.

XX This sequence is encoded by a human prostate tumour specific gene, and

CC can be used in the method of the invention. The method is for detecting

CC prostate cancer comprises contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC this protein sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

CC (Updated on 25-MAR-2003 to correct PR field.)

XX SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 19; Length 553;  
Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLRRHKAQLLLVNLITFGLEVCLAAGITYVPPLLEVGVEEKFMTVLGIG 60  
DB 1 MVQRLWVSRLRRHKAQLLLVNLITFGLEVCLAAGITYVPPLLEVGVEEKFMTVLGIG 60  
QY 61 PVLGLVCVPLIGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
DB 61 PVLGLVCVPLIGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPHCRQAYSVYAFMISLGGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPHCRQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240  
DB 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240  
QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTLRLLFVAELCSWMLMTFTLYTDFVGEGL 300  
DB 241 CCPCARLAFRNLGALLPRHLQCCMRPTLRLLFVAELCSWMLMTFTLYTDFVGEGL 300  
QY 301 YQGVPRABPTGTEARRHYDEGVRMGSLGLFQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
DB 301 YQGVPRABPTGTEARRHYDEGVRMGSLGLFQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGSGLLPPPPALCGASACDVSVRVVVGSPTEA 480  
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGSGLLPPPPALCGASACDVSVRVVVGSPTEA 480  
QY 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540

Db 481 RVVPGRGICLDLAIDSAFLLSQVAPSLFMGSIQVLSQVTAVMVSAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553  
RESULT 3  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
AC AAB28527;  
XX  
DT 07-FEB-2001 (first entry)  
XX  
XX Protein encoded by human breast tumour cDNA clone P501S.  
DE Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine.  
XX  
XX Homo sapiens.  
XX WO200061756-A2.  
PN 19-OCT-2000.  
XX  
XX 10-APR-2000; 2000WO-US09688.  
XX  
PR 09-APR-1999; 99US-0288950.  
PR 02-JUL-1999; 99US-0346327.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Reed SG, Xu J, Dillon DC;  
XX  
DR WPI; 2000-638568/61.  
DR N-PSDB; AAC79473.  
XX  
PT A novel isolated polypeptide comprising an immunogenic portion of a  
PT breast cancer protein useful in the detection and treatment of breast  
PT cancer -  
XX  
XX Claim 2; Page 92-93; 95pp; English.  
XX  
XX The present sequence is encoded by a cDNA sequence which was isolated  
XX from a breast tumour cDNA library. It is provided in a specification  
XX relating to compounds for immunotherapy and diagnosis of breast cancer.  
XX Breast tumour antigens and the polynucleotides that encode them may be  
XX used in the production of a pharmaceutical composition to be used in the  
XX treatment of breast cancer. Proliferated T cells and incubated antigen  
XX presenting cells are also required. The polypeptides and polynucleotides  
XX may also be used to produce a vaccine.  
XX  
SQ Sequence 553 AA;  
Query Match 100.0%; Score 2861; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVQRLWVSRLLRHKAKQLLLVNLITFGLEVLAAAGITYVPPLLLEVGVEEKFMTWVLGIG 60  
Db 1 MVQRLWVSRLLRHKAKQLLLVNLITFGLEVLAAAGITYVPPLLLEVGVEEKFMTWVLGIG 60  
QY 61 PVGLVCVPLLSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Db 61 PVGLVCVPLLSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
QY 121 ELALLILGVLLDFCCQVCTPTEALLSDFRDPDHCROAYSFYAFWISLGGCLGYLLPA 180  
Db 121 ELALLILGVLLDFCCQVCTPTEALLSDFRDPDHCROAYSFYAFWISLGGCLGYLLPA 180  
QY 181 IDWDTSAAPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPSLSPH 240

Db 181 IDWDTSAAPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPSLSPH 240  
QY 241 CCPCRLARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTFLFYDFVGEGL 300  
Db 241 CCPCRLARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTFLFYDFVGEGL 300  
QY 301 YQGVPAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFRTAVYLASVA 360  
Db 301 YQGVPAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFRTAVYLASVA 360  
QY 361 APFVAAGATCLSHSVAVVTASAAALTGFTPSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 APFVAAGATCLSHSVAVVTASAAALTGFTPSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFNGHVAGGSGLLPPPPALCGASACDVSRRVVGEPTEA 480  
Db 421 ASSEDSLMTSFLPGPKGAPFNGHVAGGSGLLPPPPALCGASACDVSRRVVGEPTEA 480  
QY 481 RVVPGRGICLDLAIDSAFLLSQVAPSLFMGSIQVLSQVTAVMVSAAGLGLVAIFYATQ 540  
Db 481 RVVPGRGICLDLAIDSAFLLSQVAPSLFMGSIQVLSQVTAVMVSAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553  
RESULT 4  
AAY82002  
ID AAY82002 standard; Protein; 553 AA.  
XX  
AC AAY82002;  
XX  
DT 13-JUN-2000 (first entry)  
XX  
DE Human immunogenic prostate tumour protein sequence SEQ ID NO:113.  
XX  
KW Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KW immunogenic; cytostatic; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200004149-A2.  
XX  
PD 27-JAN-2000.  
XX  
PF 14-JUL-1999; 99WO-US15838.  
XX  
PR 14-JUL-1998; 98US-0115453.  
PR 14-JUL-1998; 98US-0116134.  
PR 23-SEP-1998; 98US-0159812.  
PR 23-SEP-1998; 98US-0159822.  
PR 15-JAN-1999; 99US-0232149.  
PR 15-JAN-1999; 99US-0232880.  
PR 09-APR-1999; 99US-0288946.  
XX  
XX (CORI-) CORIXA CORP.  
PA  
PI Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;  
XX  
XX WPI; 2000-171268/15.  
DR  
XX  
PT New polypeptide useful for treating and diagnosing prostate cancer  
PT comprises an immunogenic portion of prostate tumor protein -  
XX  
XX Claim 3; Page 138-139; 263pp; English.  
XX  
XX The present invention describes isolated polypeptides, comprising an  
XX immunogenic portion of a prostate tumor protein (PTP). The polypeptides  
XX and polynucleotides encoding them have cytostatic activity and can be  
XX used in vaccines and in gene therapy. The polypeptides and  
XX polynucleotides encoding them, antigen presenting cells which express

CC the polypeptides, antibodies against the polypeptides and vaccines  
 CC comprising them can be used for inhibiting the development of prostate  
 CC cancer in a patient. The polypeptides can be used to generate antibodies  
 CC or anti-idiotypic antibodies for passive immuno therapy. A portion of  
 CC the polynucleotides encoding the polypeptides can be used as a probe or  
 CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and  
 CC AA82000 to AA82020 represent sequences used in the exemplification of  
 CC the present invention.

XX SQ Sequence 553 AA;  
 Query Match 100.0%; Score 2861; DB 21; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSLRLRRKKAQLLLVNLTLFGLEVCLAAGITYVPPLLEVGVEKFMVTLGIG 60  
 DB 1 MVQRLWVSLRLRRKKAQLLLVNLTLFGLEVCLAAGITYVPPLLEVGVEKFMVTLGIG 60  
 QY 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWTSALAPYLGTQECLFGLTLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWTSALAPYLGTQECLFGLTLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLARLAFNLGALLPRHLQCCRMPTLRRLFVAELCSWALMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLARLAFNLGALLPRHLQCCRMPTLRRLFVAELCSWALMTFTLYTDFVGEGL 300  
 QY 301 YQGVPRAEPTGEARRHYDEGVNMGSLGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQGVPRAEPTGEARRHYDEGVNMGSLGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAIQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAIQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSLMTSFLPGPKGAPFNGHVAGAGSGLLPPPPALCGASACDVSVRVVVGTEA 480  
 DB 421 ASSEDSLMTSFLPGPKGAPFNGHVAGAGSGLLPPPPALCGASACDVSVRVVVGTEA 480  
 QY 481 RVVPGRGICLDLALDLSAFLLSOVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540  
 DB 481 RVVPGRGICLDLALDLSAFLLSOVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540  
 QY 541 VVFDKSLAKYSA 553  
 DB 541 VVFDKSLAKYSA 553

RESULT 5  
 AAU69763  
 ID AAU69763 standard; Protein; 553 AA.  
 AC AAU69763;  
 XX  
 XX 30-JAN-2002 (first entry)  
 XX Human prostate cDNA encoded protein #3.  
 XX Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.  
 XX Homo sapiens.  
 XX W0200173032-A2.  
 XX 04-OCT-2001.

XX PF 27-MAR-2001; 2001WO-US09919.  
 XX PR 27-MAR-2000; 2000US-0536857.  
 PR 09-MAY-2000; 2000US-0568100.  
 PR 12-MAY-2000; 2000US-0570737.  
 PR 13-JUN-2000; 2000US-0593793.  
 PR 27-JUN-2000; 2000US-0605783.  
 PR 10-AUG-2000; 2000US-0636215.  
 PR 29-AUG-2000; 2000US-0651236.  
 PR 06-SEP-2000; 2000US-0657279.  
 PR 02-OCT-2000; 2000US-0679426.  
 PR 10-OCT-2000; 2000US-0685166.  
 XX (CORI-) CORIXA CORP.  
 PA Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedwick TS, Carter D;  
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;  
 XX WPI; 2001-639232/73.  
 DR N-PSDB; AAS63557.  
 XX New human prostate-specific polypeptides and polynucleotides useful for  
 PT the diagnosis and treatment of cancer, especially prostate cancer -  
 PT Claim 2; Page 269-270; 579pp; English.  
 XX The invention relates to isolated prostate-specific  
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,  
 CC antibodies raised against the polypeptides for antigenic epitopes  
 CC derived from them) and antigen-presenting cells expressing the  
 CC polypeptides. The antibodies are useful for detecting the presence of  
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
 CC the antigen-presenting cells are useful for stimulating and/or expanding  
 CC T cells specific for a tumour protein, and for inhibiting the development  
 CC of cancer especially prostate cancer. Compositions comprising the  
 CC polynucleotide and/or polypeptide are useful for stimulating an immune  
 CC response, and for treating cancer. The oligonucleotide is useful for  
 CC detecting cancer. The present sequence is a prostate specific  
 CC polypeptide of the invention.

XX SQ Sequence 553 AA;  
 Query Match 100.0%; Score 2861; DB 22; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSLRLRRKKAQLLLVNLTLFGLEVCLAAGITYVPPLLEVGVEKFMVTLGIG 60  
 DB 1 MVQRLWVSLRLRRKKAQLLLVNLTLFGLEVCLAAGITYVPPLLEVGVEKFMVTLGIG 60  
 QY 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWTSALAPYLGTQECLFGLTLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWTSALAPYLGTQECLFGLTLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLARLAFNLGALLPRHLQCCRMPTLRRLFVAELCSWALMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLARLAFNLGALLPRHLQCCRMPTLRRLFVAELCSWALMTFTLYTDFVGEGL 300  
 QY 301 YQGVPRAEPTGEARRHYDEGVNMGSLGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQGVPRAEPTGEARRHYDEGVNMGSLGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAIQILPYTLASLYHREKQVFLPKYRGDTGG 420

```
Db 361 APPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGSLPPPPALCGASACDVSVRVVVGPEPTA 480
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGSLPPPPALCGASACDVSVRVVVGPEPTA 480
QY 481 RVVPGRGICLDLAIDLDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAIFYATQ 540
Db 481 RVVPGRGICLDLAIDLDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAIFYATQ 540
QY 541 VVFDKSDIAKYSA 553
Db 541 VVFDKSDIAKYSA 553

RESULT 6
AAU04961
ID AAU04961 standard; Protein; 553 AA.
AC AAU04961;
DT 24-OCT-2001 (first entry)
DE Human prostate tumour protein L1-12.
KW Human; prostate tumour protein; prostate cancer.
XX Homo sapiens.
OS
XX
PN US6262245-B1.
PD 17-JUL-2001.
PF 25-FEB-1998; 98US-0030607.
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
PR 09-FEB-1998; 98US-0020956.
XX
PA (CORI-) CORIXA CORP.
XX
XX Xu J, Dillon DC;
PI WPI; 2001-440862/47.
DR N-PSDB; AAS10108.
XX
XX Novel polynucleotide encoding polypeptide comprising a portion of
PT prostate tumour protein useful for inhibiting development of prostate
PT cancer or for treating prostate cancer in a patient
XX
PS Example 1; Column 125-127; 105pp; English.
XX
XX The sequence is a partial prostate tumour protein, encoded by a prostate
CC tumour specific cDNA. The DNA is useful for inhibiting the development
CC of prostate cancer or for treating prostate cancer in a patient.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 2.2e-269;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLHRKQAQLLVNLLTFGLVCLAGITVPPPLLLLVGVVEEKFMTNVLGIG 60
Db 1 MVQRLWVSRLLHRKQAQLLVNLLTFGLVCLAGITVPPPLLLLVGVVEEKFMTNVLGIG 60
QY 61 PVGLVCPVLLGSADHWGRGRRPFIWALSGLLSLFLIPAGWLAGLCPDPRPL 120
Db 61 PVGLVCPVLLGSADHWGRGRRPFIWALSGLLSLFLIPAGWLAGLCPDPRPL 120
QY 121 ELALLILGVGLDFCGQVCFTPLEALLSDLFRDPDHCQAYSVYAFMISLGGCLGYLLPA 180
```

```
Db 121 ELALLILGVGLDFCGQVCFTPLEALLSDLFRDPDHCQAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSAALPYLGTQEBCLFGLLTLLITCVAAATLLVAEEAALGPTPEAGLSAPLSLSPH 240
Db 181 IDWDTSAALPYLGTQEBCLFGLLTLLITCVAAATLLVAEEAALGPTPEAGLSAPLSLSPH 240
QY 241 CCPCRLAFLRNGLGALLPRLHQLCCRMPTLRLRLEVAELCSMMALMTFTLFTYDFVGEGL 300
Db 241 CCPCRLAFLRNGLGALLPRLHQLCCRMPTLRLRLEVAELCSMMALMTFTLFTYDFVGEGL 300
QY 301 YQGVPRAPGTEARHYDEGVRMGSLGLFLOCAISLVFSLVMDRLVQRPRTGTRAVLASVA 360
Db 301 YQGVPRAPGTEARHYDEGVRMGSLGLFLOCAISLVFSLVMDRLVQRPRTGTRAVLASVA 360
QY 361 APPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db 361 APPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGSLPPPPALCGASACDVSVRVVVGPEPTA 480
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGSLPPPPALCGASACDVSVRVVVGPEPTA 480
QY 481 RVVPGRGICLDLAIDLDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAIFYATQ 540
Db 481 RVVPGRGICLDLAIDLDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAIFYATQ 540
QY 541 VVFDKSDIAKYSA 553
Db 541 VVFDKSDIAKYSA 553

RESULT 7
AAU01117
ID AAU01117 standard; Protein; 553 AA.
AC AAU01117;
DT 04-OCT-2001 (first entry)
DE Human prostate-specific amino acid sequence L1-12.
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX Homo sapiens.
XX WO200151633-A2.
XX 19-JUL-2001.
XX 16-JAN-2001; 2001WO-US01574.
XX 14-JAN-2000; 2000US-0483672.
XX (CORI-) CORIXA CORP.
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX WPI; 2001-425873/45.
XX
XX New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines
XX
PS Claim 2; Page 267-268; 543pp; English.
XX
XX The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
```

CC (I) and the antibodies are also used in the detection of cancer in a  
 CC patient. The cancer that is diagnosed or treated is particularly  
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or  
 CC (I) can be used for monitoring the progression of cancer in a patient.  
 CC (I) and (II) can also be used to improve diagnostic and therapeutic  
 CC methods for prostate cancer. They can indicate the level of metastasis  
 CC as well as the prostate volume. AAH93357 to AAH93944 and AAH01115 to  
 CC AAH01318 represent polynucleotide and amino acid sequences used in the  
 CC exemplification of the present invention.

XX Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 22; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2,2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLLVLLTFTGLEVCVLAAGITVPPPLLELVGVBEKFTMTVLGIG 60  
 DB 1 MVQRLWVSRLLRHRKAQLLLVLLTFTGLEVCVLAAGITVPPPLLELVGVBEKFTMTVLGIG 60  
 QY 61 PVLGLVCPVLLGSASDHWRGRRPFTWALSGLILLSLFLIPRAGWLAGLLCPDPRLL 120  
 DB 61 PVLGLVCPVLLGSASDHWRGRRPFTWALSGLILLSLFLIPRAGWLAGLLCPDPRLL 120  
 QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWDTSAALPYLGTQECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWDTSAALPYLGTQECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCARLAFRNLGALLPRHLCCRMPTLRLLFVAELCSWMLMTFTLFTYDFVGEGL 300  
 DB 241 CCPCARLAFRNLGALLPRHLCCRMPTLRLLFVAELCSWMLMTFTLFTYDFVGEGL 300  
 QY 301 YQGVPRAEPTGTEARRHYDEGVGMGSLGLFQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQGVPRAEPTGTEARRHYDEGVGMGSLGLFQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAALQILPYTLASLYHREKQVFLPKYRGDTG 420  
 DB 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAALQILPYTLASLYHREKQVFLPKYRGDTG 420  
 QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
 DB 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIIVQLSQSVTAYMWSAAGLGLVLIYFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIIVQLSQSVTAYMWSAAGLGLVLIYFATQ 540  
 QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

## RESULT 8

AAAG99002  
 ID AAAG99002 standard; Protein; 553 AA.

XX AC AAAG99002;

XX DT 25-SEP-2001 (first entry)

XX DE Human prostate-specific amino acid sequence L1-12/P501S.

XX KW Human; prostate cancer; therapy; diagnosis; cat eye syndrome;  
 KW chromosome 22q11.2; prostate-specific protein; chromosome 1;  
 KW prostate specific antigen; PSA.

XX OS Homo sapiens.

PN WO200134802-A2.

XX PD 17-MAY-2001.

XX PF 09-NOV-2000; 2000WO-US30904.

XX PR 12-NOV-1999; 99US-0439313.

XX PR 18-NOV-1999; 99US-0443686.

XX PA (CORI-) CORIXA CORP.

XX PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;

XX PI Kalos WD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;

XX DR WPI; 2001-308785/32.

XX Isolated polypeptide comprising at least an immunogenic portion of a  
 PT prostate-specific protein, useful in the diagnosis and therapy of  
 PT prostate cancer -

PS Claim 3; Page 167-168; 325pp; English.

XX The present invention describes an isolated polypeptide (P1) comprising  
 CC at least an immunogenic portion of a prostate-specific protein, or its  
 CC variant. Also described are polynucleotides (N1) encoding (P1) and  
 CC (N1) have cytostatic activity and can be used in vaccine production.  
 CC The polypeptides, nucleic acids and antibodies from the present  
 CC invention are useful in the diagnosis and therapy of prostate cancer.  
 CC Prostate specific genes P704P, P712P, P774P, P775P and B305D are located  
 CC in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome  
 CC region. Prostate specific antigen (PSA) P501S was located on  
 CC chromosome 1. AAH84671 to AAH85143 and AAAG99000 to AAAG99077 represent  
 CC polynucleotide and polypeptide sequences used in the exemplification  
 CC of the present invention.

XX Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 22; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2,2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLLVLLTFTGLEVCVLAAGITVPPPLLELVGVBEKFTMTVLGIG 60

DB 1 MVQRLWVSRLLRHRKAQLLLVLLTFTGLEVCVLAAGITVPPPLLELVGVBEKFTMTVLGIG 60

QY 61 PVLGLVCPVLLGSASDHWRGRRPFTWALSGLILLSLFLIPRAGWLAGLLCPDPRLL 120

DB 61 PVLGLVCPVLLGSASDHWRGRRPFTWALSGLILLSLFLIPRAGWLAGLLCPDPRLL 120

QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180

DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240

DB 181 IDWDTSAALPYLGTQECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240

QY 241 CCPCARLAFRNLGALLPRHLCCRMPTLRLLFVAELCSWMLMTFTLFTYDFVGEGL 300

DB 241 CCPCARLAFRNLGALLPRHLCCRMPTLRLLFVAELCSWMLMTFTLFTYDFVGEGL 300

QY 301 YQGVPRAEPTGTEARRHYDEGVGMGSLGLFQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360

DB 301 YQGVPRAEPTGTEARRHYDEGVGMGSLGLFQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360

QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAALQILPYTLASLYHREKQVFLPKYRGDTG 420

DB 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAALQILPYTLASLYHREKQVFLPKYRGDTG 420

QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

DB 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIYFATQ 540  
 DB |||||  
 QY 481 RVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIYFATQ 540  
 DB |||||  
 QY 541 VVFDKSLAKYSA 553  
 DB |||||  
 QY 541 VVFDKSLAKYSA 553  
 DB |||||

RESULT 9  
 AAG62150  
 ID AAG62150 standard; Protein; 553 AA.  
 XX  
 AC AAG62150;  
 XX  
 DT 06-JUL-2001 (first entry)  
 XX  
 DE Human P501S inventive antigen SEQ ID NO: 333.  
 XX  
 KW Human; mouse; immunotherapy; cancer; leukaemia; WTI; Wilm's tumour gene;  
 KW chromosome 11p13; zinc finger transcription factor.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200125273-A2.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 04-OCT-2000; 2000WO-US27465.  
 XX  
 PR 04-OCT-1999; 99US-0157459.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Xu J, Cheever MA, Reed SG;  
 XX  
 DR WPI; 2001-328324/34.  
 XX  
 XX Polypeptide comprising part of the Wilm's Tumour gene product sequence is  
 PT used in the diagnosis and treatment of malignant diseases e.g. leukemia  
 PT and cancer associated with WTI  
 XX  
 PS Disclosure; Page 212-213; 228pp; English.  
 XX  
 CC The present invention describes compositions comprising peptides derived  
 CC from the Wilm's tumour protein WTI and methods for their use in treating  
 CC malignant diseases. Peptides derived from both the murine and human WTI  
 CC proteins are provided. The human WTI gene is found on chromosome 11p13,  
 CC and the protein was shown to be a zinc finger transcription factor. The  
 CC -immunogenic peptides of the invention are particularly useful in the  
 CC diagnosis and treatment of cancer and leukaemia. The present sequence is  
 CC a polypeptide described in the exemplification of the invention.  
 XX  
 SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 22; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLRHRKAQLLLVNLFTFGLVCLAAAGITVPPLLLEVGVVEEKFMTMVLGIG 60  
 DB |||||  
 QY 61 PVLLGVCPVLGASDHWGRGRRRPFIALSIGILLSFLIPRAGWLAGLCPDRPL 120  
 DB |||||  
 QY 61 PVLLGVCPVLGASDHWGRGRRRPFIALSIGILLSFLIPRAGWLAGLCPDRPL 120  
 DB |||||  
 QY 121 ELALLILGVGLDFCGQVCFTPLLEALLSDLPDPDHCQAVSVVAFMISLGGCLGYLLPA 180  
 DB |||||  
 QY 121 ELALLILGVGLDFCGQVCFTPLLEALLSDLPDPDHCQAVSVVAFMISLGGCLGYLLPA 180  
 DB |||||  
 QY 181 IDWTSALAPYLGTQOECLFGLLTILFTCVAAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB |||||

DB 181 IDWTSALAPYLGTQOECLFGLLTILFTCVAAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY |||||  
 DB 241 CCPCARLAFFNLGALLPRLHQLCCRMPTLRLRFVAELCSMMALMTTFLFYDFVGEGL 300  
 DB |||||  
 DB 241 CCPCARLAFFNLGALLPRLHQLCCRMPTLRLRFVAELCSMMALMTTFLFYDFVGEGL 300  
 DB |||||  
 QY 301 YQGVPRAPPGTEARRHYDEGVRMGSIGLFLQCAISLVSFLVMDRLVQRFGTTRAVYLASYA 360  
 DB |||||  
 DB 301 YQGVPRAPPGTEARRHYDEGVRMGSIGLFLQCAISLVSFLVMDRLVQRFGTTRAVYLASYA 360  
 DB |||||  
 QY 361 APVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB |||||  
 DB 361 APVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB |||||  
 QY 421 ASSSDSLMTSFLPGPKGAPFNGHVAGGSGLLPPPPALCGASACDYSVRVVVGEPTEA 480  
 DB |||||  
 DB 421 ASSSDSLMTSFLPGPKGAPFNGHVAGGSGLLPPPPALCGASACDYSVRVVVGEPTEA 480  
 DB |||||  
 QY 481 RVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIYFATQ 540  
 DB |||||  
 QY 541 VVFDKSLAKYSA 553  
 DB |||||  
 DB 541 VVFDKSLAKYSA 553  
 DB |||||

RESULT 10  
 AAB74800  
 ID AAB74800 standard; Protein; 553 AA.  
 XX  
 AC AAB74800;  
 XX  
 DT 14-JUN-2001 (first entry)  
 XX  
 DE Prostate tumour antigen predicted amino acid sequence for L1-12.  
 XX  
 KW Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;  
 KW prostate cancer; immunogenic; cytostatic; vaccine.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200125272-A2.  
 XX  
 PD 12-APR-2001.  
 XX  
 PF 04-OCT-2000; 2000WO-US27464.  
 XX  
 PR 04-OCT-1999; 99US-0157455.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Xu J, Skeiky YAW, Reed SG, Cheever MA;  
 XX  
 DR WPI; 2001-245062/25.  
 XX  
 PR N-PSDB; AAH02530.  
 XX  
 PT Prostate specific protein and its encoding polynucleotide, useful for  
 PT the treatment and diagnosis of prostate cancer  
 XX  
 PS Claim 3; Page 157-158; 276pp; English.  
 XX  
 CC The present invention describes an isolated polypeptide (I) comprising  
 CC at least an immunogenic portion of a prostate tumour antigen protein or  
 CC its variant. (I) have cytostatic activity and can be used in vaccine  
 CC production. (II), prostate tumour antigen polynucleotides, an antigen  
 CC presenting cell (APC e.g. a dendritic cell) that expresses (I), and a  
 CC pharmaceutical composition containing (I) are useful for inhibiting the  
 CC development of cancer in a patient. Antibodies specific for prostate  
 CC specific proteins and oligonucleotides that hybridise to a  
 CC polynucleotide that encodes a prostate specific protein are useful  
 CC for detecting the presence or absence of a cancer or monitoring the  
 CC progression the progression of a cancer, especially prostate cancer.





XX ABG76665;  
XX  
XX  
XX 05-NOV-2002 (first entry)  
XX  
XX Prostate tumour protein #3.  
XX  
XX Human; prostate tumour; immunotherapy; prostate cancer.  
XX  
XX Homo sapiens.  
XX  
XX OS  
XX US2002081580-A1:  
XX  
XX 27-JUN-2002.  
XX  
XX 25-FEB-1998; 98US-0030606.  
XX  
XX 25-FEB-1997; 97US-0806596.  
XX  
XX 01-AUG-1997; 97US-0904809.  
XX  
XX 09-FEB-1998; 98US-0020747.  
XX  
XX (XUJ/) XU J.  
XX  
XX (DILL/) DILLON D C.  
XX  
XX  
XX Xu J, Dillon DC;  
XX  
XX WPI; 2002-607662/65.  
XX  
XX  
XX Detecting prostate cancer comprises contacting a sample with an agent  
XX capable of binding to a polypeptide with an immunogenic portion of a  
XX prostate protein, oligonucleotide primers or a probe specific for DNA  
XX encoding the polypeptide  
XX  
XX Example 1; Page 65-67; 111pp; English.  
XX  
XX The invention relates to a method of detecting prostate cancer by  
XX contacting a biological sample from a patient with: (a) a binding agent  
XX that binds to a polypeptide having an immunogenic portion of a prostate  
XX protein or its variant; (b) 2 oligonucleotide primers, where 1 of the  
XX oligonucleotides is specific for a DNA encoding the polypeptide of (a);  
XX or (c) an oligonucleotide probe specific for a DNA molecule encoding the  
XX polypeptide of (a). The method and polypeptides are useful for  
XX diagnosing, treating, particularly by immunotherapy, monitoring the  
XX progression, and inhibiting the development of prostate cancer in a  
XX patient. The polypeptides may be used to generate antibodies useful for  
XX the diagnosis and monitoring of prostate cancer. ABG76663-ABG76669  
XX represent human prostate tumour protein sequences of the invention.  
XX  
XX Sequence 553 AA;  
XX  
XX Query Match  
XX Best Local Similarity 100.0%; Score 2861; DB 23; Length 553;  
XX Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
XX 1 MYQRLWVSRLLRHRKAQLLNLLTFGLVCLAGITVPPLLLEVGVEEKFMTVLGIG 60  
XX  
XX 1 MYQRLWVSRLLRHRKAQLLNLLTFGLVCLAGITVPPLLLEVGVEEKFMTVLGIG 60  
XX  
XX 61 PVLGLVVCVPLIGSADHWGRYGRRRPTIWLISGILLSLFLIPRAGWLAGLLCPDRPL 120  
XX  
XX 61 PVLGLVVCVPLIGSADHWGRYGRRRPTIWLISGILLSLFLIPRAGWLAGLLCPDRPL 120  
XX  
XX 121 ELALLILGVLLDFCGQVCFPLEALLSDLPRDPHCRQAVSVAFMISLGGCLGYLLPA 180  
XX  
XX 121 ELALLILGVLLDFCGQVCFPLEALLSDLPRDPHCRQAVSVAFMISLGGCLGYLLPA 180  
XX  
XX 181 IDWDTLSALAPYLGTECECLFGLLTILFTCVAAATLLVAEEAALGTEPAEGLSAPLSPH 240  
XX  
XX 181 IDWDTLSALAPYLGTECECLFGLLTILFTCVAAATLLVAEEAALGTEPAEGLSAPLSPH 240  
XX  
XX 241 CCPCARLAFRNGLGALLPRLHQLCCMRPTRLRLRFLVABLCSWMAIMTTLFTYDFVSGEL 300  
XX  
XX 241 CCPCARLAFRNGLGALLPRLHQLCCMRPTRLRLRFLVABLCSWMAIMTTLFTYDFVSGEL 300

QY 301 YQGVPRAEPTGTEARRHYDEGVMSIGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
DB 301 YQGVPRAEPTGTEARRHYDEGVMSIGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSAALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSAALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGGSLPPPPALCGASACDVSRVVVVGEPTEA 480  
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGGSLPPPPALCGASACDVSRVVVVGEPTEA 480  
QY 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMYSAAGLGLVAIFYATQ 540  
DB 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMYSAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553  
RESULT 13  
ABB77575  
ID ABB77575 standard; Protein; 553 AA.  
XX  
XX AC ABB77575;  
XX  
XX DT 30-AUG-2002 (first entry)  
XX  
XX DE Human mast cell related splice variant protein MC14 SEQ ID NO 13.  
XX  
XX KW Human; mast cell; MC; antiallergic; antiinflammatory; antiasthmatic;  
XX vasotropic; dermatological; allergy; hypersensitivity; rhinitis; asthma.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200246389-A2.  
XX  
XX PD 13-JUN-2002.  
XX  
XX PF 07-DEC-2001; 2001WO-US46180.  
XX  
XX PR 08-DEC-2000; 2000US-251835P.  
XX  
XX PR 14-MAR-2001; 2001US-275479P.  
XX  
XX PR 28-MAR-2001; 2001US-279115P.  
XX  
XX PR 02-APR-2001; 2001US-280143P.  
XX  
XX PA (UNIO ) UCB SA.  
XX  
XX PI Nocka K, Pirozzi G, Einstein R;  
XX  
XX XX WPI; 2002-508560/54.  
XX  
XX DR N-PSDB; AEN81324.  
XX  
XX PT Novel isolated nucleic acids that are differentially expressed in mast  
XX cells in patients with allergic hypersensitivity, encoding proteins  
XX associated with mast cell regranulation and allergic hypersensitivity  
XX  
XX PS Claim 31; Page 117-119; 119pp; English.  
XX  
XX CC The invention relates to isolated nucleic acid (AEN81319-AEN81324),  
XX corresponding to genes differentially expressed in mast cells following  
XX activation or in patients with allergic hypersensitivity disease, (I)  
XX that encodes proteins (ABB77569-ABB77575) (II) or a protein fragment of  
XX (II) if at least 6 amino acids. (II) is useful for identifying binding  
XX partners. (I) or (II) is useful for diagnosing or treating a disease  
XX state (e.g. allergic hypersensitivity, seasonal rhinitis, asthma,  
XX urticaria or atopic dermatitis or mastocytosis) in a subject which  
XX involves determining the level of expression of (I) or (II). A computer  
XX system, comprising a database containing information identifying the  
XX expression level in a tissue or at least one mast cell of (I), is useful

CC for presenting information to identify the relative expression level of  
 CC (I) (II) is used as a marker to detect, diagnose or identify an allergic  
 CC response in a patient. The protein can also serve as a target that  
 CC modulate gene expression or activity and as an antigen to raise  
 CC polyclonal or monoclonal antibodies. (II) is useful for identifying  
 CC agents that modulate expression of the protein or agents, such as  
 CC agonists or antagonists. The agonists or antagonists are useful for  
 CC modulating biological activity and function of (II) and thus are useful  
 CC for alleviating disease conditions such as allergic hypersensitivity,  
 CC seasonal rhinitis, asthma, urticaria, atopic dermatitis or mastocytosis.  
 XX  
 SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 23; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLVWSRLRRHKAQLLLVNLITFGLVCLAAAGITYVPPLLEVGVEEKFMTWLGIG 60  
 DB 1 MVQRLVWSRLRRHKAQLLLVNLITFGLVCLAAAGITYVPPLLEVGVEEKFMTWLGIG 60

QY 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
 DB 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120

QY 121 ELALLITLGVCLDFCGOVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLITLGVCLDFCGOVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240  
 DB 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240

QY 241 CCPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSWMLMTFTLYTDFVGEGL 300  
 DB 241 CCPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSWMLMTFTLYTDFVGEGL 300

QY 301 YQGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRECTRAVYLASVA 360  
 DB 301 YQGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRECTRAVYLASVA 360

QY 361 APFVAAGATCLSHSVAVVTASALTGFTFSALQILPVTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 APFVAAGATCLSHSVAVVTASALTGFTFSALQILPVTLASLYHREKQVFLPKYRGDTGG 420

QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGSGSLPPPPALCGASACDVSVRVVVGEPTEA 480  
 DB 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGSGSLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQTAYMVSAGIGLVAIYFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQTAYMVSAGIGLVAIYFATQ 540

QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

RESULT 14  
 ABG61900  
 ID ABG61900 standard; Protein; 553 AA.  
 XX  
 AC ABG61900;  
 XX  
 DT 15-AUG-2002 (first entry)  
 XX  
 DE Prostate cancer-associated protein #101.  
 XX  
 KW Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.  
 XX  
 OS Mammalia.  
 XX  
 PN WO200230268-A2.

XX 18-APR-2002.  
 PD  
 XX 12-OCT-2001; 2001WO-US32045.  
 PF  
 XX 13-OCT-2000; 2000US-0687576.  
 PR 08-DEC-2000; 2000US-0733288.  
 PR 08-DEC-2000; 2000US-0733742.  
 PR 24-JAN-2001; 2001US-263957P.  
 PR 16-MAR-2001; 2001US-276791P.  
 PR 16-MAR-2001; 2001US-276888P.  
 PR 06-APR-2001; 2001US-281922P.  
 PR 24-APR-2001; 2001US-286214P.  
 PR 30-APR-2001; 2001US-0847046.  
 PR 04-MAY-2001; 2001US-288589P.  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 PA  
 XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;  
 PI  
 XX WPI: 2002-471335/50.  
 XX N-PSDB; ABK92217.  
 DR  
 XX  
 DR Detecting a prostate cancer-associated transcript in a cell in a  
 PT patient, useful for diagnosing prostate cancer (PC) or screening  
 PT modulators of PC, by determining if prostate cancer-associated genes  
 PT are expressed in a prostate tissue -  
 XX  
 PS Claim 27; Page 386; 436pp; English.  
 XX  
 CC The present invention relates to methods of detecting a prostate  
 CC cancer-associated transcript in a cell from a patient. The method  
 CC comprises contacting a biological sample from the patient with  
 CC prostate cancer-associated polynucleotides (designated PC genes) that  
 CC selectively hybridise to a sequence that is at least 80% identical  
 CC to them. The prostate cancer-associated polynucleotide sequences  
 CC are differentially expressed in prostate tumour tissue or in  
 CC prostate cancer and are derived from the tissues of various  
 CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
 CC The methods of the invention are useful for diagnosing and treating  
 CC prostate cancer in mammals. The prostate cancer-associated genes are  
 CC useful for diagnosing or treating prostate cancer, as well as for  
 CC identifying modulators of prostate cancer or agents that inhibit  
 CC prostate cancer. The nucleic acid sequences are particularly useful  
 CC in gene therapy, as a vaccine or in antisense applications.  
 CC ABG61800-ABG61944 represent prostate cancer-associated proteins.  
 XX  
 SQ Sequence 553 AA;

Query Match 100.0%; Score 2861; DB 23; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 2.2e-269;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLVWSRLRRHKAQLLLVNLITFGLVCLAAAGITYVPPLLEVGVEEKFMTWLGIG 60  
 DB 1 MVQRLVWSRLRRHKAQLLLVNLITFGLVCLAAAGITYVPPLLEVGVEEKFMTWLGIG 60

QY 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
 DB 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120

QY 121 ELALLITLGVCLDFCGOVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLITLGVCLDFCGOVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240  
 DB 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAGLSAPLSLPH 240

QY 241 CCPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSWMLMTFTLYTDFVGEGL 300  
 DB 241 CCPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSWMLMTFTLYTDFVGEGL 300



Search completed: December 3, 2003, 17:19:18  
Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 3, 2003, 17:18:21 ; Search time 36 Seconds  
(without alignments)  
2856.916 Million cell updates/sec

Title: US-09-593-793a-113

Perfect score: 2861

Sequence: 1 MVQRLVWSRLHRKAQLL.....AIYFATQVVDKSLAKYSA 553

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2861	100.0	553	9	US-09-745-288-101
2	2861	100.0	553	9	US-09-838-785-2
3	2861	100.0	553	9	US-09-759-143-113
4	2861	100.0	553	9	US-09-780-669-113
5	2861	100.0	553	9	US-09-030-606-113
6	2861	100.0	553	9	US-09-822-827-113
7	2861	100.0	553	9	US-09-115-453-113
8	2861	100.0	553	10	US-09-232-880-113
9	2861	100.0	553	10	US-09-895-793-113
10	2861	100.0	553	10	US-09-895-814-113
11	2861	100.0	553	12	US-10-144-678A-113
12	2861	100.0	553	12	US-10-005-907-13
13	2861	100.0	553	12	US-10-294-025-113
14	2861	100.0	553	14	US-10-012-896-113
15	2861	100.0	553	15	US-10-010-940-113

16	2861	100.0	1079	9	US-09-822-827-947	Sequence 947, App
17	2861	100.0	1079	10	US-09-895-793-947	Sequence 947, App
18	2601	90.9	710	15	US-10-296-770-4	Sequence 4, Appli
19	1696	59.3	359	9	US-09-822-827-974	Sequence 974, App
20	1696	59.3	359	10	US-09-895-793-974	Sequence 974, App
21	1677.5	58.6	305	15	US-10-296-770-5	Sequence 5, Appli
22	1517	53.0	305	12	US-10-144-678A-1029	Sequence 1029, Ap
23	1517	53.0	305	12	US-10-294-025-1029	Sequence 1029, Ap
24	1417.5	49.5	371	9	US-09-759-143-708	Sequence 708, App
25	1417.5	49.5	371	9	US-09-780-669-708	Sequence 708, App
26	1417.5	49.5	371	9	US-09-822-827-708	Sequence 708, App
27	1417.5	49.5	371	10	US-09-895-793-708	Sequence 708, App
28	1417.5	49.5	371	10	US-09-895-814-708	Sequence 708, App
29	1417.5	49.5	371	12	US-10-144-678A-708	Sequence 708, App
30	1417.5	49.5	371	12	US-10-294-025-708	Sequence 708, App
31	1417.5	49.5	371	14	US-10-012-896-708	Sequence 708, App
32	1403.5	49.1	400	9	US-09-759-143-852	Sequence 852, App
33	1403.5	49.1	400	9	US-09-780-669-852	Sequence 852, App
34	1403.5	49.1	400	9	US-09-822-827-852	Sequence 852, App
35	1403.5	49.1	400	10	US-09-895-793-852	Sequence 852, App
36	1403.5	49.1	400	10	US-09-895-814-852	Sequence 852, App
37	1403.5	49.1	400	12	US-10-144-678A-852	Sequence 852, App
38	1403.5	49.1	400	12	US-10-294-025-852	Sequence 852, App
39	1403.5	49.1	400	14	US-10-012-896-852	Sequence 852, App
40	1197	41.8	246	12	US-10-144-678A-1028	Sequence 1028, Ap
41	1197	41.8	246	12	US-10-294-025-1028	Sequence 1028, Ap
42	1014.5	35.5	326	12	US-10-005-907-5	Sequence 5, Appli
43	947	33.1	355	12	US-10-144-678A-1011	Sequence 1011, Ap
44	947	33.1	355	12	US-10-294-025-1011	Sequence 1011, Ap
45	947	33.1	355	14	US-10-012-896-1011	Sequence 1011, Ap

## ALIGNMENTS

## RESULT 1

US-09-745-288-101  
; Sequence 101, Application US/09745288  
; Patent No. US20010018058A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, David C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.446D1  
; CURRENT APPLICATION NUMBER: US/09/745,288  
; CURRENT FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 101  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-745-288-101

Query Match	100.0%;	Score 2861;	DB 9;	Length 553;
Best Local Similarity	100.0%;	Pred. No. 1.9e-245;		
Matches 553;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MYQRLVWSRLHRKAQLLNNLLTFGLVCLAAAGTTPVPLLLLEVGVEEKMTWLIGT	60	
Db	1	MYQRLVWSRLHRKAQLLNNLLTFGLVCLAAAGTTPVPLLLLEVGVEEKMTWLIGT	60	
QY	61	PVLGLVCVPLLGASDHWRGRRPFIWALSGLLSLFLIPRAGMLAGLLCPDPRPL	120	
Db	61	PVLGLVCVPLLGASDHWRGRRPFIWALSGLLSLFLIPRAGMLAGLLCPDPRPL	120	
QY	121	ELALLILGVLDFGCGVCFPLEALLSDLRDPDHCQAYSVYAFMISLGCIGYLLPA	180	
Db	121	ELALLILGVLDFGCGVCFPLEALLSDLRDPDHCQAYSVYAFMISLGCIGYLLPA	180	
QY	181	IDWDTSALAPYLGTECECLFGLLTILFTCVAAATLVAAEALGTFEPAEGLSATSLSPH	240	

Db 181 IDWDTSAALPYLGTQBECLFGLLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
Qy 241 CCPCARLAFRNLGALLPRHLQCCRMPTLRLLFVAELCSMMALMTFTLYTDFVGEGL 300  
Db 241 CCPCARLAFRNLGALLPRHLQCCRMPTLRLLFVAELCSMMALMTFTLYTDFVGEGL 300  
Qy 301 YQGVPRAPGTEARRHYDEGVMSGLGLFLOCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
Db 301 YQGVPRAPGTEARRHYDEGVMSGLGLFLOCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
Qy 361 AFPVAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 AFPVAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Qy 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Qy 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIIVQLSQSVTAYMWSAAGLGLVAYIFATQ 540  
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIIVQLSQSVTAYMWSAAGLGLVAYIFATQ 540  
Qy 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

## RESULT 2

US-09-838-785-2  
; Sequence 2, Application US/09838785  
; Patent No. US20020009455A1  
; GENERAL INFORMATION:  
; APPLICANT: Lau, Ted  
; APPLICANT: Lin, Rick  
; APPLICANT: Parkes, Debbie  
; APPLICANT: Parry, Gordon  
; APPLICANT: Schneider, Douglas  
; APPLICANT: Steinbrecher, Renate  
; APPLICANT: Van Heut, Pam T  
; APPLICANT: Wu, John  
; TITLE OF INVENTION: DNA Encoding a No. US20020009455A1el PROST 03  
; FILE REFERENCE: 51831AUSM1  
; CURRENT APPLICATION NUMBER: US/09/838,785  
; CURRENT FILING DATE: 2001-04-20  
; PRIOR FILING DATE: 2000-04-27  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-838-785-2

Query Match 100.0%; Score 2861; DB 9; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MVORLWVSRLRRHKAQLLLVNLTFGLVCLAAAGTTPVPLLEVGVBKEKFTMWLGIG 60  
Db 1 MVORLWVSRLRRHKAQLLLVNLTFGLVCLAAAGTTPVPLLEVGVBKEKFTMWLGIG 60  
Qy 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Db 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Qy 121 ELALLIIGVLLDPCGOVCTTPEALLSDLFROPDHCQRAYSVYAFMISLGGCLGYLLPA 180  
Db 121 ELALLIIGVLLDPCGOVCTTPEALLSDLFROPDHCQRAYSVYAFMISLGGCLGYLLPA 180  
Qy 181 IDWDTSAALPYLGTQBECLFGLLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240

Db 181 IDWDTSAALPYLGTQBECLFGLLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
Qy 241 CCPCARLAFRNLGALLPRHLQCCRMPTLRLLFVAELCSMMALMTFTLYTDFVGEGL 300  
Db 241 CCPCARLAFRNLGALLPRHLQCCRMPTLRLLFVAELCSMMALMTFTLYTDFVGEGL 300  
Qy 301 YQGVPRAPGTEARRHYDEGVMSGLGLFLOCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
Db 301 YQGVPRAPGTEARRHYDEGVMSGLGLFLOCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
Qy 361 AFPVAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 AFPVAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Qy 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Qy 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIIVQLSQSVTAYMWSAAGLGLVAYIFATQ 540  
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIIVQLSQSVTAYMWSAAGLGLVAYIFATQ 540  
Qy 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

## RESULT 3

US-09-759-143-113  
; Sequence 113, Application US/09759143  
; Patent No. US2002002248A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yuqui  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Carter, Darrick  
; APPLICANT: Li, Samuel  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.427C23  
; CURRENT APPLICATION NUMBER: US/09/759,143  
; CURRENT FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 934  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 113  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-759-143-113

Query Match 100.0%; Score 2861; DB 9; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MVORLWVSRLRRHKAQLLLVNLTFGLVCLAAAGTTPVPLLEVGVBKEKFTMWLGIG 60  
Db 1 MVORLWVSRLRRHKAQLLLVNLTFGLVCLAAAGTTPVPLLEVGVBKEKFTMWLGIG 60  
Qy 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Db 61 PVGLVLCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120

```

QY 121 ELALLILGVLGDLDFCCQVCFPTLEALLSDFRDPDHCROAYSVYAFWISLGGCLGYLLPA 180
DB 121 ELALLILGVLGDLDFCCQVCFPTLEALLSDFRDPDHCROAYSVYAFWISLGGCLGYLLPA 180
QY 181 IDWDTSALAPYLGTQBECLFGLLTFLTCVAATLLVABEAAALGPTPEAGLSAPLSPH 240
DB 181 IDWDTSALAPYLGTQBECLFGLLTFLTCVAATLLVABEAAALGPTPEAGLSAPLSPH 240
QY 241 CCPCARLAFRNLGALLPRLHQLCCMRPRTLRLFVAELCSWMLMTFTFLFYDFVGEGL 300
DB 241 CCPCARLAFRNLGALLPRLHQLCCMRPRTLRLFVAELCSWMLMTFTFLFYDFVGEGL 300
QY 301 YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
DB 301 YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
QY 361 APVAAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
DB 361 APVAAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
QY 481 RVVPGRGICLDLALDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540
DB 481 RVVPGRGICLDLALDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540
QY 541 VVFDKSDLAKYSA 553
DB 541 VVFDKSDLAKYSA 553

```

RESULT 4

```

US-09-780-669-113
; Sequence 113, Application US/09780669
; Patent No. US20020051977A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C24
; CURRENT APPLICATION NUMBER: US/09/780,669
; CURRENT FILING DATE: 2001-02-09
; NUMBER OF SEQ ID NOS: 943
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-780-669-113
Query Match 100.0%; Score 2861; DB 9; Length 553;

```

```

Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MYQRLWVSLRHRKAQLLLVNLLTFGLVCLAAAGITVYPPDLLLVGVBEKFWMTWLIG 60
DB 1 MYQRLWVSLRHRKAQLLLVNLLTFGLVCLAAAGITVYPPDLLLVGVBEKFWMTWLIG 60
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFTIWLISLGLTSLFLIPRAGWLAGLLCPDPRPL 120
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFTIWLISLGLTSLFLIPRAGWLAGLLCPDPRPL 120
QY 121 ELALLILGVLGDLDFCCQVCFPTLEALLSDFRDPDHCROAYSVYAFWISLGGCLGYLLPA 180
DB 121 ELALLILGVLGDLDFCCQVCFPTLEALLSDFRDPDHCROAYSVYAFWISLGGCLGYLLPA 180
QY 181 IDWDTSALAPYLGTQBECLFGLLTFLTCVAATLLVABEAAALGPTPEAGLSAPLSPH 240
DB 181 IDWDTSALAPYLGTQBECLFGLLTFLTCVAATLLVABEAAALGPTPEAGLSAPLSPH 240
QY 241 CCPCARLAFRNLGALLPRLHQLCCMRPRTLRLFVAELCSWMLMTFTFLFYDFVGEGL 300
DB 241 CCPCARLAFRNLGALLPRLHQLCCMRPRTLRLFVAELCSWMLMTFTFLFYDFVGEGL 300
QY 301 YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
DB 301 YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
QY 361 APVAAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
DB 361 APVAAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
QY 481 RVVPGRGICLDLALDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540
DB 481 RVVPGRGICLDLALDLSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540
QY 541 VVFDKSDLAKYSA 553
DB 541 VVFDKSDLAKYSA 553

```

RESULT 5

```

US-09-030-606-113
; Sequence 113, Application US/09030606
; Patent No. US20020081580A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHODS FC
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030,606
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.428C3

```

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 622-4900  
; TELEFAX: (206) 682-6031  
; INFORMATION FOR SEQ ID NO: 113:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 553 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; US-09-030-606-113

Query Match 100.0%; Score 2861; DB 9; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVQRLWVSRLLRRKKAQLLVNLLTFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60  
Db 1 MVQRLWVSRLLRRKKAQLLVNLLTFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60  
QY 61 PVGLVVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Db 61 PVGLVVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
QY 121 ELALITLGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
Db 121 ELALITLGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTSALAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLSPH 240  
Db 181 IDWDTSALAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLSPH 240  
QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSWMLMTFTLFYDFVGEGL 300  
Db 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSWMLMTFTLFYDFVGEGL 300  
QY 301 YQGVPAEPGTEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
Db 301 YQGVPAEPGTEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
QY 361 APFVAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 APFVAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIFYATQ 540  
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

## RESULT 6

US-09-822-827-113  
; Sequence 113, Application US/09822827  
; Patent No. US20020081680A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; FILE REFERENCE: 210121.534C1  
; CURRENT APPLICATION NUMBER: US/09/822, 827  
; CURRENT FILING DATE: 2001-03-28  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 113

; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapien  
; US-09-822-827-113

Query Match 100.0%; Score 2861; DB 9; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRRKKAQLLVNLLTFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60  
Db 1 MVQRLWVSRLLRRKKAQLLVNLLTFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60  
QY 61 PVGLVVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Db 61 PVGLVVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
QY 121 ELALITLGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
Db 121 ELALITLGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTSALAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLSPH 240  
Db 181 IDWDTSALAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLSPH 240  
QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSWMLMTFTLFYDFVGEGL 300  
Db 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSWMLMTFTLFYDFVGEGL 300  
QY 301 YQGVPAEPGTEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
Db 301 YQGVPAEPGTEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
QY 361 APFVAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 APFVAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIFYATQ 540  
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

## RESULT 7

US-09-115-453-113  
; Sequence 113, Application US/09115453B  
; Patent No. US20020090372A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND  
; TITLE OF INVENTION: METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.427C4  
; CURRENT APPLICATION NUMBER: US/09/115, 453B  
; CURRENT FILING DATE: 1998-07-14  
; NUMBER OF SEQ ID NOS: 228  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 113  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapien  
; US-09-115-453-113

Query Match 100.0%; Score 2861; DB 9; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 MVORLWVSRLLRHRKAQLLVNLITFGLVCLAAAGITYVPPILLLEVGVBEKFTMWLGIG 60  
DB 1 MVORLWVSRLLRHRKAQLLVNLITFGLVCLAAAGITYVPPILLLEVGVBEKFTMWLGIG 60  
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPDHCRCQAYSVYAFMISLGCGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPDHCRCQAYSVYAFMISLGCGCLGYLLPA 180  
QY 181 IDWDTLSALAPYLGTQBECLFGLLTLLIFLTCTVAATLLVAEEAALGPTPEAGLSAPSLSPH 240  
DB 181 IDWDTLSALAPYLGTQBECLFGLLTLLIFLTCTVAATLLVAEEAALGPTPEAGLSAPSLSPH 240  
QY 241 CPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
DB 241 CPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
QY 301 YQGVPRAEPTGTEARRHYDEGRVMSGLFLQCAISLVESLVMDRLVQRFCTRAVYLASVA 360  
DB 301 YQGVPRAEPTGTEARRHYDEGRVMSGLFLQCAISLVESLVMDRLVQRFCTRAVYLASVA 360  
QY 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
DB 421 ASSEDSMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
QY 481 RVVPGRGICLDLAILDSAPLLSOVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
DB 481 RVVPGRGICLDLAILDSAPLLSOVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553

RESULT 8  
US-09-232-880-113  
; Sequence 113, Application US/09232880  
; Publication No. US20020182596A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Mitcham, Jennifer Lynn  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF  
; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.428C6  
; CURRENT APPLICATION NUMBER: US/09/232,880  
; CURRENT FILING DATE: 1999-01-15  
; NUMBER OF SEQ ID NOS: 338  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 113  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-232-880-113

Query Match 100.0%; Score 2861; DB 10; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVORLWVSRLLRHRKAQLLVNLITFGLVCLAAAGITYVPPILLLEVGVBEKFTMWLGIG 60  
DB 1 MVORLWVSRLLRHRKAQLLVNLITFGLVCLAAAGITYVPPILLLEVGVBEKFTMWLGIG 60  
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120

DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPDHCRCQAYSVYAFMISLGCGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGOVCFPLEALLSDFRDPDHCRCQAYSVYAFMISLGCGCLGYLLPA 180  
QY 181 IDWDTLSALAPYLGTQBECLFGLLTLLIFLTCTVAATLLVAEEAALGPTPEAGLSAPSLSPH 240  
DB 181 IDWDTLSALAPYLGTQBECLFGLLTLLIFLTCTVAATLLVAEEAALGPTPEAGLSAPSLSPH 240  
QY 241 CPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
DB 241 CPCRARLAFRNIGALLPRLHQLCCMRPTRLRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
QY 301 YQGVPRAEPTGTEARRHYDEGRVMSGLFLQCAISLVESLVMDRLVQRFCTRAVYLASVA 360  
DB 301 YQGVPRAEPTGTEARRHYDEGRVMSGLFLQCAISLVESLVMDRLVQRFCTRAVYLASVA 360  
QY 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
DB 421 ASSEDSMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
QY 481 RVVPGRGICLDLAILDSAPLLSOVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
DB 481 RVVPGRGICLDLAILDSAPLLSOVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553

RESULT 9  
US-09-895-793-113  
; Sequence 113, Application US/09895793  
; Publication No. US20020192763A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan L.  
; APPLICANT: Jiang, Yugu  
; APPLICANT: Kalos, Michael D.  
; APPLICANT: Retter, Marc W.  
; APPLICANT: Stolk, John A.  
; APPLICANT: Day, Craig H.  
; APPLICANT: Vedwick, Thomas S.  
; APPLICANT: Carter, Darriek  
; APPLICANT: Li, Samuel X.  
; APPLICANT: Wang, Aijun  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Hepler, William T.  
; APPLICANT: Henderson, Robert A.  
; APPLICANT: Hural, John  
; APPLICANT: McNeill, Patricia D.  
; APPLICANT: Houghton, Raymond L.  
; APPLICANT: Vinals de Baesols, Carlota  
; APPLICANT: Foy, Teresa  
; APPLICANT: Fanger, Gary R.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; FILE REFERENCE: 210121.534C2  
; CURRENT APPLICATION NUMBER: US/09/895,793  
; CURRENT FILING DATE: 2001-06-29  
; NUMBER OF SEQ ID NOS: 982  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 113  
; LENGTH: 553  
; TYPE: PRT  
; ORGANISM: Homo sapien

US-09-895-793-113

```

Query Match      100.0%; Score 2861; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLVNLTLFTGLEVCLAAGITYVPPLLLEVGVEEKEFTMWVLGIG 60
Db 1 MVQRLWVSRLLRHRKAQLLVNLTLFTGLEVCLAAGITYVPPLLLEVGVEEKEFTMWVLGIG 60

QY 61 PVGLVVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120
Db 61 PVGLVVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVGLDFCCQVCFPLEALLSDLPDPCRCQAYSVYAFMISLGGCLGYLLPA 180
Db 121 ELALLILGVGLDFCCQVCFPLEALLSDLPDPCRCQAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQECFLGILLTLFTCVAAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240
Db 181 IDWDTSAALPYLGTQECFLGILLTLFTCVAAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240

QY 241 CCPCARLAFRNLGALLPRLHQLCCRMPTLRLLFVAELCSMMALMTFTLFYDFVGEGL 300
Db 241 CCPCARLAFRNLGALLPRLHQLCCRMPTLRLLFVAELCSMMALMTFTLFYDFVGEGL 300

QY 301 YQGVPRAPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLSPA 360
Db 301 YQGVPRAPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLSPA 360

QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

QY 421 ASSEDSLMTSFLPGPKPGAPPFNHGVGAGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
Db 421 ASSEDSLMTSFLPGPKPGAPPFNHGVGAGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVOLQSQTAYMVSAAGLGLVAIFYATQ 540
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVOLQSQTAYMVSAAGLGLVAIFYATQ 540

QY 541 VVFDKSDLAKYSA 553
Db 541 VVFDKSDLAKYSA 553

```

RESULT 10

```

US-09-895-814-113
; Sequence 113, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Gary R.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa

```

```

; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; CURRENT FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
; US-09-895-814-113

Query Match      100.0%; Score 2861; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLVNLTLFTGLEVCLAAGITYVPPLLLEVGVEEKEFTMWVLGIG 60
Db 1 MVQRLWVSRLLRHRKAQLLVNLTLFTGLEVCLAAGITYVPPLLLEVGVEEKEFTMWVLGIG 60

QY 61 PVGLVVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120
Db 61 PVGLVVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVGLDFCCQVCFPLEALLSDLPDPCRCQAYSVYAFMISLGGCLGYLLPA 180
Db 121 ELALLILGVGLDFCCQVCFPLEALLSDLPDPCRCQAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQECFLGILLTLFTCVAAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240
Db 181 IDWDTSAALPYLGTQECFLGILLTLFTCVAAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240

QY 241 CCPCARLAFRNLGALLPRLHQLCCRMPTLRLLFVAELCSMMALMTFTLFYDFVGEGL 300
Db 241 CCPCARLAFRNLGALLPRLHQLCCRMPTLRLLFVAELCSMMALMTFTLFYDFVGEGL 300

QY 301 YQGVPRAPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLSPA 360
Db 301 YQGVPRAPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLSPA 360

QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

QY 421 ASSEDSLMTSFLPGPKPGAPPFNHGVGAGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
Db 421 ASSEDSLMTSFLPGPKPGAPPFNHGVGAGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVOLQSQTAYMVSAAGLGLVAIFYATQ 540
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVOLQSQTAYMVSAAGLGLVAIFYATQ 540

QY 541 VVFDKSDLAKYSA 553
Db 541 VVFDKSDLAKYSA 553

RESULT 11
US-10-144-678A-113
; Sequence 113, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.

```

APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A. W.  
APPLICANT: Hepler, William T.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals y de Bassols, Carlota  
APPLICANT: Foy, Teresa M.  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Deng, Ta  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C28  
CURRENT APPLICATION NUMBER: US/10/144, 678A  
CURRENT FILING DATE: 2002-08-12  
NUMBER OF SEQ ID NOS: 1033  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-144-678A-113

Query Match 100.0%; Score 2861; DB 12; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MVORLWVSRLLRHRKAQLLLVNLITFGLVCLAAAGITVPPLLLLEVGVEEKFMTVLGIG 60  
Db 1 MVORLWVSRLLRHRKAQLLLVNLITFGLVCLAAAGITVPPLLLLEVGVEEKFMTVLGIG 60  
Qy 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGLAGLCCPDPRPL 120  
Db 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGLAGLCCPDPRPL 120  
Qy 121 ELALLILGVLLDFCGQVCFPLEALLSDLPDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
Db 121 ELALLILGVLLDFCGQVCFPLEALLSDLPDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
Qy 181 IDWDTSAAPYLGTQECFLGLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
Db 181 IDWDTSAAPYLGTQECFLGLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
Qy 241 CCPCRARLAFRNIGALLPRLHQLCCRPRTLRLFLVAELCSMWALMTFTLYTDFVGEGL 300  
Db 241 CCPCRARLAFRNIGALLPRLHQLCCRPRTLRLFLVAELCSMWALMTFTLYTDFVGEGL 300  
Qy 301 YQGVPAEPGTARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
Db 301 YQGVPAEPGTARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
Qy 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Qy 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Db 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Qy 481 RVVPGRGICLDLAILDSAPLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
Db 481 RVVPGRGICLDLAILDSAPLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
Qy 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

## RESULT 13

US-10-294-025-113  
Sequence 113, Application US/10294025  
Publication No. US20030185830A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun

RESULT 12  
US-10-005-907-13  
Sequence 13, Application US/10005907  
Publication No. US20030166881A1  
GENERAL INFORMATION:  
APPLICANT: Union Chimique Belge, S.A.  
APPLICANT: No. US20030166881A1ka, Karl  
APPLICANT: Pirozzi, Gregory  
APPLICANT: Einstein, Richard  
TITLE OF INVENTION: NOVEL GENES ASSOCIATED WITH ALLERGIC HYPERSENSITIVITY AND MAST CELL  
TITLE OF INVENTION: ACTIVATION  
FILE REFERENCE: 053529-5005  
CURRENT APPLICATION NUMBER: US/10/005, 907  
CURRENT FILING DATE: 2001-12-07  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 13  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-005-907-13

Query Match 100.0%; Score 2861; DB 12; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MVORLWVSRLLRHRKAQLLLVNLITFGLVCLAAAGITVPPLLLLEVGVEEKFMTVLGIG 60  
Db 1 MVORLWVSRLLRHRKAQLLLVNLITFGLVCLAAAGITVPPLLLLEVGVEEKFMTVLGIG 60  
Qy 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGLAGLCCPDPRPL 120  
Db 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGLAGLCCPDPRPL 120  
Qy 121 ELALLILGVLLDFCGQVCFPLEALLSDLPDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
Db 121 ELALLILGVLLDFCGQVCFPLEALLSDLPDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
Qy 181 IDWDTSAAPYLGTQECFLGLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
Db 181 IDWDTSAAPYLGTQECFLGLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
Qy 241 CCPCRARLAFRNIGALLPRLHQLCCRPRTLRLFLVAELCSMWALMTFTLYTDFVGEGL 300  
Db 241 CCPCRARLAFRNIGALLPRLHQLCCRPRTLRLFLVAELCSMWALMTFTLYTDFVGEGL 300  
Qy 301 YQGVPAEPGTARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
Db 301 YQGVPAEPGTARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
Qy 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Qy 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Db 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
Qy 481 RVVPGRGICLDLAILDSAPLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
Db 481 RVVPGRGICLDLAILDSAPLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIFYATQ 540  
Qy 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

```
; APPLICANT: Stolk, John A.
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C29
; CURRENT APPLICATION NUMBER: US/10/294,025
; CURRENT FILING DATE: 2002-11-12
; NUMBER OF SEQ ID NOS: 1038
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-294-025-113

Query Match      100.0%; Score 2861; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLELVGVVEKFMVWLIG 60
DB 1 MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLELVGVVEKFMVWLIG 60
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIFWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIFWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
QY 121 ELALLILGVGLDFCGQVCFPLEALLSDLFDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180
DB 121 ELALLILGVGLDFCGQVCFPLEALLSDLFDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSLAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
DB 181 IDWDTSLAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
QY 241 CCPCARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFYDFVGEGL 300
DB 241 CCPCARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFYDFVGEGL 300
QY 301 YQGVPRAEPTGEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFQTRAVYL 360
DB 301 YQGVPRAEPTGEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFQTRAVYL 360
QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKY 420
DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKY 420
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVG 480
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVG 480
QY 481 RVVPGRGICLDLAIDLAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAI 540
DB 481 RVVPGRGICLDLAIDLAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAI 540
QY 541 VFPDKSLAKYSA 553
DB 541 VFPDKSLAKYSA 553
```

## RESULT 14

```
US-10-012-896-113
; Sequence 113, Application US/10012896
; Publication No. US20020183251A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillion, Devin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yugu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
```

```
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; APPLICANT: Wantanabe, Yoshihiro
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C27
; CURRENT APPLICATION NUMBER: US/10/012,896
; CURRENT FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 1011
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-896-113
```

```
Query Match      100.0%; Score 2861; DB 14; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9e-245;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLELVGVVEKFMVWLIG 60
DB 1 MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLELVGVVEKFMVWLIG 60
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIFWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIFWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
QY 121 ELALLILGVGLDFCGQVCFPLEALLSDLFDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180
DB 121 ELALLILGVGLDFCGQVCFPLEALLSDLFDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSLAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
DB 181 IDWDTSLAPYLGTOBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
QY 241 CCPCARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFYDFVGEGL 300
DB 241 CCPCARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFYDFVGEGL 300
QY 301 YQGVPRAEPTGEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFQTRAVYL 360
DB 301 YQGVPRAEPTGEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFQTRAVYL 360
QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKY 420
DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKY 420
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVG 480
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVG 480
QY 481 RVVPGRGICLDLAIDLAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAI 540
DB 481 RVVPGRGICLDLAIDLAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAI 540
QY 541 VFPDKSLAKYSA 553
DB 541 VFPDKSLAKYSA 553
```

## RESULT 15

US-10-010-940-113  
; Sequence 113, Application US/10010940  
; Publication No. US20030089062A1  
; GENERAL INFORMATION:  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqui  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael  
; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark  
; APPLICANT: Solk, John  
; APPLICANT: Day, Craig

## TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427D3

; CURRENT APPLICATION NUMBER: US/10/010,940

; CURRENT FILING DATE: 2001-12-05

; NUMBER OF SEQ ID NOS: 575

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 113

; LENGTH: 553

; TYPE: PRT

; ORGANISM: Homo sapien

US-10-010-940-113

Query Match 100.0%; Score 2861; DB 15; Length 553;  
Best Local Similarity 100.0%; Pred. No. 1.9e-245;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVQRLWVSRLRRHKAQILLNLTFTGLEVCCLAAGITVVPPLLELVGVBEKFTMTVLGIG	60
Db	1	MVQRLWVSRLRRHKAQILLNLTFTGLEVCCLAAGITVVPPLLELVGVBEKFTMTVLGIG	60
Qy	61	PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCCPDPRPL	120
Db	61	PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCCPDPRPL	120
Qy	121	ELALLILGVLLDFCGQVCFPLEALLSDLFRDPDHCRCQAYSVYAFMISLGGCLGYLLPA	180
Db	121	ELALLILGVLLDFCGQVCFPLEALLSDLFRDPDHCRCQAYSVYAFMISLGGCLGYLLPA	180
Qy	181	IDWDTALAPYLGTOEBCFLGLTLIFLTCTVAATLLVAEEAALGPTPEAEGLSAPLSPH	240
Db	181	IDWDTALAPYLGTOEBCFLGLTLIFLTCTVAATLLVAEEAALGPTPEAEGLSAPLSPH	240
Qy	241	CCPCRARLAFNGLALLPRHLQCCRPRTLRLFLVAELCSWMAALMTFTLYTDFVGEGL	300
Db	241	CCPCRARLAFNGLALLPRHLQCCRPRTLRLFLVAELCSWMAALMTFTLYTDFVGEGL	300
Qy	301	YQGVPRAEPTGEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVORFGTRAVYLASVA	360
Db	301	YQGVPRAEPTGEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVORFGTRAVYLASVA	360
Qy	361	APVVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG	420
Db	361	APVVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG	420
Qy	421	ASSEDSLMTSFLCPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA	480
Db	421	ASSEDSLMTSFLCPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA	480
Qy	481	RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYVWSAAGLGLVAIYFATQ	540
Db	481	RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYVWSAAGLGLVAIYFATQ	540
Qy	541	VVPDKSDLAKYSA	553
Db	541	VVPDKSDLAKYSA	553

Search completed: December 3, 2003, 17:22:03  
Job time : 37 secs

---

**THIS PAGE BLANK (USPTO)**

---

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:17:41 ; Search time 21 Seconds  
(without alignments)  
2532.442 Million cell updates/sec

Title: US-09-593-793a-113  
Perfect score: 2861  
Sequence: 1 MVQRLWVSRLRHRKAQLL.....AIYFATQVDFKDLAKYSA 553

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 76.\*  
1: Pir1.\*  
2: Pir2.\*  
3: Pir3.\*  
4: Pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	347.5	12.1	515	2 T14340	sucrose-proton tra
2	337.5	11.8	523	2 T12198	sucrose transport
3	330	11.5	516	2 JQ2389	sucrose transport
4	328.5	11.5	594	2 G84441	probable sucrose-p
5	324.5	11.3	525	2 S28052	sucrose transport
6	323.5	11.3	512	2 S38196	sucrose transport
7	323.5	11.3	512	2 G86360	probable sucrose-p
8	317.5	11.1	507	2 S48789	sucrose transport
9	311	10.9	533	2 S43142	sucrose transport
10	305.5	10.7	501	2 T14339	sucrose-proton tra
11	297.5	10.4	428	2 S48788	sucrose transport
12	294	10.3	510	2 S38657	sucrose transport
13	294	10.3	523	2 S51114	sucrose-proton sym
14	292.5	10.2	512	2 E96741	probable sucrose t
15	292	10.2	537	2 T02982	probable sucrose c
16	286	10.0	474	2 A86234	hypothetical prote
17	281.5	9.8	491	2 E96691	probable sucrose-p
18	279.5	9.8	513	2 S38197	sucrose transport
19	276.5	9.7	492	2 A84520	probable sucrose-p
20	273	9.5	503	2 S52377	sucrose transport
21	235.5	8.2	553	2 T38541	probable sucrose c
22	191	6.7	452	2 F75217	hypothetical prote
23	187.5	6.6	541	2 B87532	transporter, proba
24	142	5.0	544	2 S75696	melibiose carrier
25	139.5	4.9	454	2 A75444	hypothetical prote
26	138.5	4.8	430	2 E75217	transporter PAB217
27	136	4.8	418	2 B87536	membrane protein,
28	133	4.6	389	2 G83413	probable MFS trans
29	128.5	4.5	422	2 G83503	probable MFS trans

probable permease  
resistance protein  
probable integral  
tetracycline resis  
drug efflux protei  
sugar transporter,  
tetracycline resis  
hypothetical prote  
tetracycline-efflu  
teta protein - pse  
MFS permease [drug  
probable efflux pr  
hypothetical metab  
multidrug-efflux t  
puromycin resistan  
hypothetical prote

ALIGNMENTS

RESULT 1

T14340  
sucrose-proton transport protein - carrot  
N;Alternate names: sucrose/H+ symporter protein  
C;Species: Daucus carota (carrot)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000  
C;Accession: T14340  
R;Shakya, R.; Sturm, A.  
Plant Physiol. 118, 1473-1480, 1998  
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from carrot  
A;Reference number: Z17991; MUID:99063785; PMID:9847123  
A;Accession: T14340  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-515 <SHA>  
A;Cross-references: EMBL:Y16768; NID:G2969883; PIDN:CAA76369.1; PID:G2969884  
A;Experimental source: cultivar Namtaise; root  
C;Genetics:  
A;Note: SUT2  
C;Superfamily: common tobacco sucrose transport protein

Query Match	12.1%	Score 347.5;	DB 2;	Length 515;
Best Local Similarity	25.9%;	Pred. No. 1e-18;		
Matches 131;	Conservative 80;	Mismatches 215;	Indels 79;	Gaps 17;
QY	17	QLLVNLNLFGLVCLAGITYVPPILLLEVVEERFMTMVLGIPVLGLVCPVLLGSAD	76	
Db	34	KLVLAAIAAGVQFGWALQSLTTPVQLLGIPIHKWAAVIMLCGPISGMLVQPIVGYS	93	
QY	77	HWGRYGRPRPTWALSIGLILSLFLIPRAGLAGI-----LCPDPRPLELALLIGV	129	
Db	94	HCQSSFGRRRPTFIAGAGCAISVILI---GPAADISYKAGDDMSKTLKPRATVTVFVIGF	150	
QY	130	GLLDFGQVCFPTLEALLSDLFR-DPDHCRQAYSVVAFMISLGGCLGY-----LL	178	
Db	151	WILDVANMLQPCFALLADLCSTGRMRSANAFYFPMVGNILGVAAGSYNNLYKLF	210	
QY	179	PAIDWDTSAIPYLGTOBECLFGLTLFLTCVAATLVAEEAALGPTPEAGLSAPSLS	238	
Db	211	PFS--KTHACDLYCANLKSCFFIISIALIIITVVALSVVRENS--GPPDDAABEEP---	263	
QY	239	PHCCPCRALAFRNIGALLPRHLQCCRNPRTLRLFVAELCSWALMTFTLYTDFVGE	298	
Db	264	----PSSGKIPV--FGELIGALKOL----PRMLLLLVTCNLNIAWFPFIFLFDTDWGR	313	
QY	299	GLYQGVPAEPTTEARRHRYDEGVMSGLFGLFQCAISLVFSLVMDRLVQRCFTRAVILAS	358	
Db	314	EYGGT--AGQ--KLYDQGVRAAGLLNSVVLGLTSIAVEYLVKRGVGVKI-LWG	366	
QY	359	VAAFFVAAGATCL-----SHSV-----AVVTASA-----ALTGTFFSALQIL	395	
Db	367	FYNFILAIGLVTVVSVKVAQHQREHSANGQLLPPSAGVKAGALSFLSILGIPLTYSI	426	

Qy 396 PYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSLFLPGKPG-----APFPNGHVGA 448  
Db 427 PFALASLYSSGSGAGQGLSLGVNLALIVPQMIVSVLAGPFDLSLFGGGLNLPAPFVGAISA 486  
Qy 449 GSGSLLP-----PPALCGASACDVS 469  
Db 487 AISGVLAIVLLPFSKDAASKLSLS 511

RESULT 2  
T12198  
sucrose transport protein - fava bean  
C;Species: Vicia faba (fava bean)  
C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 20-Jun-2000  
C;Accession: T12198  
R;Weber, H.; Borisjuk, L.; Heim, U.; Sauer, N.; Wobus, U.  
Plant Cell 9, 895-908, 1997  
A;Title: A role for sugar transporters during seed development: molecular characterization  
A;Reference number: Z17451; MUID:97355984; PMID:9212465  
A;Accession: T12198  
A;Status: preliminary; translated from GS/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-523 <WEB>  
A;Cross-references: EMBL:Z93774  
C;Genetics:  
A;Gene: suc  
C;Superfamily: common tobacco sucrose transport protein

Query Match 11.8%; Score 337.5; DB 2; Length 523;  
Best Local Similarity 25.8%; Pred. No. 5.9e-18;  
Matches 134; Conservative 76; Mismatches 185; Indels 125; Gaps 20;

Qy 17 QLLVNLTLFTGLVCLAAITVYPPLLLEVGVBEKFTWVLGVLGVLPVLLGSASD 76  
Db 36 KIMVVASIAAGVQFGWALQSLTTPYVQLLGIHTTAAAYIWLCPISGMLVQVPIGVHSD 95  
Qy 77 HWRGVRGRRPFTWALSGLTSLFLPRAGWLAGL-----LCPDPRLEALLILGV 129  
Db 96 RCTSRFRGRRPFTWALSGLTSLFLPRAGWLAGL-----GYAADLGHSGFSDQKVRPRATIGFVWGF 152  
Qy 130 GLLDPCQVCFPLEALLSDFR--DPDHCQAYSVVAFMISLGCGLYLLPAID-----182  
Db 153 WILDVANNMLOGPCRALLDGACNQKRTNANAFSFFVAVGNVLGYAGAYSKLYHVF 212  
Qy 183 --WDTSAPLYGTQECLFGLTLTFLTCVAATLV--AEAAALGPTEPA---EGLSAP 235  
Db 213 PFTKTRACNVYCANLKSCTF--LSIALTLVATLSALIYVKETALTPEKTVVTTEDGSSG 270  
Qy 236 SLSPHCCPCARL--APRNIGALLPRHLQCCRMPTRLRLFVABLCSWMALMTFTLYT 293  
Db 271 GM-----PCFGLSGAPKEL-----KRPMTLLLVTCINWTIAWPFLLFD 311  
Qy 294 DFVGEGLYQGVPRAEPTGEARRH--YDEGVRMGSGLFLQCAISLVFSLVMDRLVQRFGR 352  
Db 312 DMWGKEVI-----GGTVEGHAYDMGVREGALGLMNSVLGATSLGVDIILARGVG-G 363  
Qy 353 AVYLASVAAPFVAAAGATCLSHSVAV-----TASAAITGTF 387  
Db 364 VKRLGWVNFLL--AICLGLTLVTLTKLAQHSRQYAPGTGALGDPLPPSGIKAGALTIF 420  
Qy 388 TFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDLS-----MTS 430  
Db 421 SVLGVPILATYISPFALASIF-----SSTSGAQGLSLGVNLALIVPQMIVS 468  
Qy 431 FLPGKPG-----APFPNGHVGAAGSG-----LPPPPP 458  
Db 469 VLSGPDALPGGGLNLPFVVGVAALASGLSILLLPSPP 508

RESULT 3

JQ2389

sucrose transport protein - potato

C;Species: Solanum tuberosum (potato)  
C;Date: 28-Aug-1985 #sequence\_revision 07-Oct-1994 #text\_change 21-Jul-2000  
C;Accession: JQ2389; S40310  
R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.  
Plant Cell 5, 1591-1598, 1993  
A;Title: Potato sucrose transporter expression in minor veins indicates a role in phloem  
A;Reference number: JQ2389; MUID:94146554; PMID:8312741  
A;Accession: JQ2389  
A;Molecule type: mRNA  
A;Residues: 1-516 <RIE>  
A;Cross-references: EMBL:X69165; NID:G439293; PIDN:CAA48915.1; PID:G439294  
A;Experimental source: cv. Desiree  
A;Comment: The gene encoding for this protein is highly expressed in mature leaves.  
C;Superfamily: common tobacco sucrose transport protein  
C;Keywords: glycoprotein; transmembrane protein  
F;31-53/Domain: transmembrane #status predicted <TM1>  
F;103-122/Domain: transmembrane #status predicted <TM2>  
F;141-160/Domain: transmembrane #status predicted <TM3>  
F;180-200/Domain: transmembrane #status predicted <TM4>  
F;226-248/Domain: transmembrane #status predicted <TM5>  
F;285-304/Domain: transmembrane #status predicted <TM6>  
F;331-349/Domain: transmembrane #status predicted <TM7>  
F;366-385/Domain: transmembrane #status predicted <TM8>  
F;409-427/Domain: transmembrane #status predicted <TM9>  
F;429-448/Domain: transmembrane #status predicted <TM10>  
F;3/92/Binding site: carbohydrate (Asn) #status predicted  
Query Match 11.5%; Score 330; DB 2; Length 516;  
Best Local Similarity 24.7%; Pred. No. 2.2e-17;  
Matches 132; Conservative 93; Mismatches 197; Indels 112; Gaps 21;

Qy 4 RLWVSLRLHRAQLLVNLLTFGLVCLAAITVYPPLLLEVGVBEKFTWVLGIPVL 63  
Db 29 KLM-----KIIVASIAAGVQFGWALQSLTTPYVQLLGIHTTAAAYIWLCPIS 78  
Qy 64 GLVGVPLLSASDHWRGVRGRRPFTWALSGLTSLFLPRAGWLAGL-----LCPD 116  
Db 79 GMIVQVGVYSDNCSSRFGRRRPFTIAGALVMTAVFLI---GFAADLGHASGDTLGK 135  
Qy 117 PRPLELAILGLVGLDFCGVQVCTPLEALLSDFLFRDPD--HCRQAYSVVAFMISLGCGLG 175  
Db 136 FKPRATAVFWGVFWILDVANNMLOGPCRALLDGACNQKRTNANAFSFFVAVGNILG 195  
Qy 176 YLLPAIDW-----DTSALAPYLTQDECLF--GLTLTFLTCVAATLVAAE--AALGP 225  
Db 196 YAAGSYSHLKFVFPFSKTKACDMYCANLKSCTFIAIFLLSLTTLTALTLVRENELPEKDE 255  
Qy 226 TPAEGLSAPSLSPHCCPCARLAPRNIGALLPRHLQCCRMPTRLRLFVABLCSWMAL 285  
Db 256 QEIDEKLAGAG-----KSKVPF--FGEIFGALKEL-----PRPMWILLVTCINWTIAW 301  
Qy 286 MTFTLYTDFVGEGLYQGVPRAEPTGEARRH--YDEGVRMGSGLFLQCAISLVFSLVMDRL 345  
Db 302 FFFFLYDTDMAKEVEFG---QVGD--ARLYDLGVRAAGMGLLQSVVLGFWMSLGVFL 355  
Qy 346 VQRFGRTRAVLASVAAPFVAAAGATCLSHSVAV-----TASAAITGTF-----388  
Db 356 GKKIG--GAKRLWGLNLFVL---AICLAMTILVTKMAEKSRQHPDAGTLPAGTGPVIGAL 411  
Qy 389 --PSALQI-----LPYTLASLYHREK-----QVFLPKYRGDTGGASSED 426  
Db 412 LFLAALGIPLATFTSIPFALASIFSSNRSGGGLSLGVNLALIVPQMIVSLVGGVWDDL 471  
Qy 427 LMTSFLPGKPGAPFPNGHVGAAGSG-----LPPPPPACGASACDVSVRVVG 475  
Db 472 FGGGNLPG-----FVVGAVAAASAVLALTMLPSPA-----DAKPAVAMG 512

RESULT 4

G84441

probable sucrose-proton symporter [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)



C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 16-Feb-2001  
C/Accession: G84441  
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A/Reference number: A84420; MUID:20083487; PMID:10617197  
A/Accession: G84441  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-594 <STO>  
A/Cross-references: GB:AE002093; NID:G3461813; PIDN:AAC32907.1; GSPDB:GN00139  
C/Genetics:  
A/Gene: At2G02860  
A/Map position: 2  
C/Superfamily: common tobacco sucrose transport protein

Query Match 11.5%; Score 328.5; DB 2; Length 594;  
Best Local Similarity 24.1%; Pred. No. 3.3e-17;  
Matches 120; Conservative 72; Mismatches 174; Indels 131; Gaps 18;

QY 22 NLLTFLGVCVLAAGITY-----VPLLLEVGVEEKEFTWVLGIGPVGLVLCVPLLGS 73  
DB 59 SLVTLVLSCTVAGVQFGWALQSLTPTIQTIGISHAFSSFIWGLCGPITGLVWQPFVGI 118  
QY 74 ASDHWGRYGRRRPFTWALSIGLISLFIIPRAGWLA--GLLCPD-----PRPLE 121  
DB 119 WSDKCTSKYGRRRPFLVCSFMISTAVII-----GFSADIGYLLGDSKEHCSTFKGTRTA 175  
QY 122 LALLIGVGLLDFCGVCTPTLEALLSDLFDPDCHCRQAYSVAFMISIGGCLGYLLPAI 181  
DB 176 AVVFIIGFWLLDANNVQGPALLADL--SGPDQRNTANAVFLMMAIGNILGFSAGAS 234  
QY 182 -----DWDTSALAPYLGTQECFLG-----LLTLFLFT--CVAATLLVAEEAALGPTEP 228  
DB 235 GKQEW-----FPFL--TSRACCAACGNLKAFLAVVFLTICTLVITYPAKEIPFTSNKP 288  
QY 229 AE-GLSAPSLSPHCPCRLARAFRLN-----LSKGLHSLKNGTANGIKYRVERDTEQFNGSENEHQ 340  
DB 289 TRIQDSAPLLDD-----GALLPRHLQCCMRPTRLRLFAELCSWMALMTFTLYDFVGEGLYQVP 305  
QY 341 DETYVDGFGSVLNLTLRLPAMHSLVLIWALTWLSWFFFLFDIDDMGREVYHGD 400  
QY 306 RAEPCTEARRHYDEGRVMSGLFLQCAISLVFSLVMDLVRQFTRAVY--LASVAAPFV 364  
DB 401 TGD--SLHMELYDQGVREGALGILLNSVVLGISTFLIEPQCQRMGARVVMALSNFTVAC 458  
QY 365 AAGATCLS-----HSVAVVTASALGTFTFSALQILPYTLASLY 403  
DB 459 MAGTAVISLMSLSDDKNGIEYIMRGNETTRTAIV--FALLGFLPLAITYSVPFSVTA-- 514  
QY 404 HREKQVFLPKYRGDTGG 420  
DB 515 -----EVTDSGG 522

RESULT 5  
S28052  
sucrose transport protein - spinach  
N/Alternate names: sucrose carrier protein; sucrose permease  
C/Species: Spinacia oleracea (spinach)  
C/Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
C/Accession: S28052  
R/Riesmeier, J.W.; Willmitzer, L.; Frommer, W.B.  
EMBO J. 11, 4705-4713, 1992  
A/Title: Isolation and characterization of a sucrose carrier cDNA from spinach by functi  
A/Reference number: S28052; MUID:93099843; PMID:1464305  
A/Accession: S28052  
A/Molecule type: mRNA  
A/Residues: 1-525 <RIE>

A/Cross-references: EMBL:X67125; NID:G21318; PIDN:CRA47604.1; PID:G21319  
C/Superfamily: common tobacco sucrose transport protein  
C/Keywords: transmembrane protein

Query Match 11.3%; Score 324.5; DB 2; Length 525;  
Best Local Similarity 23.2%; Pred. No. 5.8e-17;  
Matches 129; Conservative 91; Mismatches 196; Indels 141; Gaps 20;

QY 23 LLTFLGLEVCVLAAGITY-----VPLLLEVGVEEKEFTWVLGIGPVGLVLCVPLLGS 74  
DB 35 LKUGLVLASVAGVQFGWALQSLTPTIQTIGISHAFSSFIWGLCGPITGLVWQPFVGI 94  
QY 75 SDHWGRYGRRRPFTWALSIGLISLFIIPRAGWLA--GLLCPD-----RPLELALLIL 127  
DB 95 SDRCTSRGRRRPFIAAGAAVAVAGLI--GFAADIGAAAGSDPTGNVAKPRAIAVFFV 151  
QY 128 GVGILLDFCGVCTPTLEALLSDLF--DPDCHCRQAYSVAFMISIGGCLGYLLPAID---- 182  
DB 152 GFWILLDVANNTLQGFPCRALADMAAGSQTTRYANAFSFFFMALGNIGGYAAGSYRLYT 211  
QY 183 ---WDTGALAPYLGTQECFLGLT--LIFLTCVATLLVAEEAALGPTEPAEGLSAPSL 237  
DB 212 VPPFTKTRACDVCYCANLSCFFISITLILVILALSVVKEQIYIDEIQEEDLKNRN 271  
QY 238 SPHCPCRLARAFRLNIGALLPRHLQCCMRPTRLRLFAELCSWMALMTFTLYDFV 297  
DB 272 SSGC---ARLPF--FGQLIGALKDL---PKPMLILLVLTALNWIAPFLLFDTDWNG 321  
QY 298 EGLYGVPRAPEGTEARRHYDEGRVMSGLFLQCAISLVFSLVMDLVRQF--TRAVY 355  
DB 322 KEVYGGT-----VGEGLYDQGVHAGALGLMINSVILGVMVMSLSIEGLARMVGGAKLWG 375  
QY 356 LASVAAPFAAGATCLSHSVAVVTASALTGTFTFSALQILPYTLASLYHREKQVFLPKYR 415  
DB 376 IVNIIL-----AVCLAMTV--LVTKSAE-----HFRDSSHIM---- 405  
QY 416 GDTGASSEDSIMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVG 475  
DB 406 -----GSAVPPPPPA--GVKGGALAFVILG 429  
QY 476 EPTEARV--VP-----GRGICLDLALDSAFLLSQV-----AP 506  
DB 430 IPLAITFIPFALASIFSGSGGGLSLGVLNLAIVVQPMPFVSVTSGWDAMFGGGLNP 489  
QY 507 SLFWGSIYQLQSQVYAT 523  
DB 490 APVVGAVAATAASAVLSF 506

RESULT 6  
S38196  
sucrose transport protein SUC2 - Arabidopsis thaliana  
N/Alternate names: sucrose-proton symporter SUC2  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000  
C/Accession: S38196; T00773  
R/Sauer, N.K.  
submitted to the EMBL Data Library, October 1993  
A/Description: SUC1 and SUC2: two sucrose transporters from Arabidopsis thaliana.  
A/Reference number: S38196  
A/Accession: S38196  
A/Molecule type: mRNA  
A/Residues: 1-512 <SAU>  
A/Cross-references: EMBL:X75382; NID:G407091; PID:G407092  
R/Vyotskaya, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Oji, O.; Kwan, A.; Liu, S.; Li,  
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel, N.  
submitted to the EMBL Data Library, July 1998  
A/Description: Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete sequence.  
A/Reference number: Z14202  
A/Accession: T00773  
A/Status: translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-490, 'A', 492-512 <VYS>

A;Cross-references: EMBL:AC003979; NID:g3172156; GSPDB:GN00059; ATSP:T22J1  
A;Experimental source: cultivar Columbia  
C;Genetics:

A;Gene: SUC2; ATSP:T22J18.12

A;Map position: 1

A;Introns: 419/3; 441/1; 455/3

C;Superfamily: common tobacco sucrose transport protein

C;Keywords: sugar transport

Query Match 11.3%; Score 323.5; DB 2; Length 512;  
Best Local Similarity 24.1%; Pred. No. 6.7e-17;  
Matches 123; Conservative 92; Mismatches 195; Indels 101; Gaps 18;

QY 14 RKAQLLLNLLTFGLVCLAAAGITVYPPILLLEVGVEEKFTMWLIGIPVGLVCPVLLGS 73  
DB 28 RLKRTISVSSIAAGVQFGWALQSLTTPYVQLLIGIPHKWASLWLCGIPSGMLVQVPIVGY 87  
QY 74 ASDHWRGRRGRRPFFIAGAGLVTVAVFLI---GYAADIGHSGMDQLDKPKTKTAIPA 144  
DB 88 HSDRCTSRFGRRPFFIAGAGLVTVAVFLI---GYAADIGHSGMDQLDKPKTKTAIPA 144  
QY 127 LGVGLDFCGQVCFPTPEALLSDFR-DPHCRQAVSVYAFMISLGGCLGY-----LL 178  
DB 145 LGFWILDVANNTLQGPCRAFLADLSAGNAKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204  
QY 179 PAIDWD-TSALAPYLCTQBECLGLTLTLTCTVAATLLVABEALGPTPEAGLSAPSL 237  
DB 205 KVPFTWTSKCDLYCANLTKTCFPLSITLL-LIVTFVSLCYVKEKPTPEPTADGKA---- 259  
QY 238 SPHCCPCRARLAFNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTFLFYDFVG 297  
DB 260 -----SNVFP--FGEIFGAFKEL-----KRPMMMLLIVTALNWIAPFLLFDTDWNG 305  
QY 298 EGYQGVPRAPGTEARRHYDEGRVMSGLGLFQCAISLVFSLVMDRLVORFGTRAVYLA 357  
DB 306 REVYGNDSATATAASKLYNDGVRAGALGLMLNAIVLGFMSLGVEMIGRKLK-GAKRLW 364  
QY 358 SVAAPFAAGATCLSHSVAV-----VTASA-----ALTG-----FT 388  
DB 365 GIWNFIL---AICLMTVVTVTKQENHRRDHGAKTGPNGVNTAGALTTLFAILGIPQAIT 421  
QY 389 FSALQILPYTLASLYHREK-----QVFLPKYRGDTGGASSEDLSMTSFLPGP 435  
DB 422 FS-----IPFALASIFSTNSGAGQGLSLGVNLAIIVPQWVISVGGGPFDELFGGGINPA- 476  
QY 436 KPGAPPNGHVGGGSGLL-----PPPPA 459  
DB 477 -----FVLGAIAAAVSGVLGLTLVLPSPPPDA 502

## RESULT 7

G86360  
probable sucrose-proton symporter SUC2 protein T22J18.12 - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Nov-2001  
C;Accession: G86360  
R;Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A;Reference number: A86141; MUID:21016719; PMID:11130712  
A;Accession: G86360  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-512 <S>O>  
A;Cross-references: GB:AE005172; NID:g3287687; PIDN:AAC25515.1; GSPDB:GN00141  
C;Genetics:

A;Map position: 1  
C;Superfamily: common tobacco sucrose transport protein

Query Match 11.3%; Score 323.5; DB 2; Length 512;  
Best Local Similarity 24.1%; Pred. No. 6.7e-17;  
Matches 123; Conservative 92; Mismatches 195; Indels 101; Gaps 18;

QY 14 RKAQLLLNLLTFGLVCLAAAGITVYPPILLLEVGVEEKFTMWLIGIPVGLVCPVLLGS 73  
DB 28 RLKRTISVSSIAAGVQFGWALQSLTTPYVQLLIGIPHKWASLWLCGIPSGMLVQVPIVGY 87  
QY 74 ASDHWRGRRGRRPFFIAGAGLVTVAVFLI---GYAADIGHSGMDQLDKPKTKTAIPA 144  
DB 88 HSDRCTSRFGRRPFFIAGAGLVTVAVFLI---GYAADIGHSGMDQLDKPKTKTAIPA 144  
QY 127 LGVGLDFCGQVCFPTPEALLSDFR-DPHCRQAVSVYAFMISLGGCLGY-----LL 178  
DB 145 LGFWILDVANNTLQGPCRAFLADLSAGNAKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204  
QY 179 PAIDWD-TSALAPYLCTQBECLGLTLTLTCTVAATLLVABEALGPTPEAGLSAPSL 237  
DB 205 KVPFTWTSKCDLYCANLTKTCFPLSITLL-LIVTFVSLCYVKEKPTPEPTADGKA---- 259  
QY 238 SPHCCPCRARLAFNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTFLFYDFVG 297  
DB 260 -----SNVFP--FGEIFGAFKEL-----KRPMMMLLIVTALNWIAPFLLFDTDWNG 305  
QY 298 EGYQGVPRAPGTEARRHYDEGRVMSGLGLFQCAISLVFSLVMDRLVORFGTRAVYLA 357  
DB 306 REVYGNDSATATAASKLYNDGVRAGALGLMLNAIVLGFMSLGVEMIGRKLK-GAKRLW 364  
QY 358 SVAAPFAAGATCLSHSVAV-----VTASA-----ALTG-----FT 388  
DB 365 GIWNFIL---AICLMTVVTVTKQENHRRDHGAKTGPNGVNTAGALTTLFAILGIPQAIT 421  
QY 389 FSALQILPYTLASLYHREK-----QVFLPKYRGDTGGASSEDLSMTSFLPGP 435  
DB 422 FS-----IPFALASIFSTNSGAGQGLSLGVNLAIIVPQWVISVGGGPFDELFGGGINPA- 476  
QY 436 KPGAPPNGHVGGGSGLL-----PPPPA 459  
DB 477 -----FVLGAIAAAVSGVLGLTLVLPSPPPDA 502

## RESULT 8

S48789  
sucrose transport protein - common tobacco  
C;Species: Nicotiana tabacum (common tobacco)  
C;Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 26-May-2000  
C;Accession: S48789  
R;Buerkle, X.Y.Z.; Frommer, W.B.  
submitted to the EMBL Data Library, October 1994  
A;Reference number: S48789  
A;Accession: S48789  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-507 <S>UE>  
A;Cross-references: EMBL:X82276; NID:g575350; PID:g575351  
C;Superfamily: common tobacco sucrose transport protein

Query Match 11.1%; Score 317.5; DB 2; Length 507;  
Best Local Similarity 24.3%; Pred. No. 1.9e-16;  
Matches 125; Conservative 87; Mismatches 203; Indels 99; Gaps 18;

QY 4 RLWVSLRLRRKRAQLLLNLLTFGLVCLAAAGITVYPPILLLEVGVEEKFTMWLIGIPVGL 63  
DB 23 KLM-----KIIMVASIAAGVQFGWALQSLTTPYVQLLIGIPHKFASFWLWCGPIS 72  
QY 64 GLVCPVLLGSADHWGRYGRRRPFFIAGAGLVTVAVFLI---GYAADIGHSGMDQLDKPKTKTAIPA 118  
DB 73 GMIVQPVVGYSDNCSSRRFRGRFTAAAGAAVTVIAVFLIGFAADL-GHATGDLGKGSK 131  
QY 119 PLEALLILGVLLDFCGVCFPTPEALLSDFR-DPHCRQAVSVYAFMISLGGCLGYLL 178

Db 132 PRAIAVVFVGVFWLIDVANNMLOGPCRALLDLGGKARMTSNAPFFFAVGNVILGYAA 191  
QY 179 PAID-----NDTSALAPVLTQBECLF-GLLTLFLTCVAATLLVAEEAALGPTPEA 229  
Db 192 GSVSRICKIPFFSKTPACDIYCANLKSCTFFIAVFLLSLTILALT--VRENELPEKDEH 249  
QY 230 EGLSAPSLSPHCCPCRARLARFRNLGALLPRHLQCCMRPTLRRLFVAELCSMMALMTFT 289  
Db 250 E-----IDEKAGARKSVFF--FGEIFGALKDL-----PRPMWILLVLTSLNIARFPFF 297  
QY 290 LFTDFVGEGLYGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFLVMDRLVQRF 349  
Db 298 LYDTDMAKEVYGG--KVGDG----RLYDLGVHAGALGLLLNSVFLGFMSLSEVFLGKKI 351  
QY 350 GTRAVVLASVAAPFAAGATCLSHSVAV-----TASALGTFTFSALQI-- 394  
Db 352 G-GVKRLWGLNLFV---AVCMALTIVLVTKMAEKSRQYDAGHTMPTSGVKIGALTFLFA 407  
QY 395 -----LPYTLASLYHREK-----QVFLPKYRGDTGGASSEDLSMTS 430  
Db 408 VLGIPLAVTFSPFALASIFSSNAGSGQGLSLGVNLAIWVQMLYSIAGGPWDDLFGGG 467  
QY 431 FLPGPKGAPFNGHVAGSG-----LPPPPA 459  
Db 468 NLPG-----FIVGAVAAAAGILALTMLPSPPA 495

RESULT 9  
S43142  
sucrose transport protein - castor bean  
N:Alternate names: sucrose carrier  
C:Species: Ricinus communis (castor bean)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 26-May-2000  
C:Accession: S43142  
R:Weig, A.; Komor, E.  
submitted to the EMBL Data Library, March 1994  
A:Description: A sucrose carrier from Ricinus communis.  
A:Reference number: S43142  
A:Accession: S43142  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-533 <WEI>  
A:Cross-references: EMBL:231561; NID:G468561; PID:G468562  
C:Superfamily: common tobacco sucrose transport protein

Query Match 10.9%; Score 311; DB 2; Length 533;  
Best Local Similarity 23.3%; Pred. No. 6.2e-16;  
Matches 134; Conservative 95; Mismatches 225; Indels 120; Gaps 19;

QY 17 QLLLVNLTFLGLEVCCLAAGITYVPPILLLEVGVEEKEFTWVLGIGPVGLVCVPLLGASD 76  
Db 37 KVVMVASIAAGIQFGWALQSLTTPVQLLGIPHTWAAFILGCGPISGMLVQPIVGYHSD 96  
QY 77 HWRGYRRRPFITWALSGLTILSLFLIPRA---GWLAG-LLCPDRPLELALLILGVGLL 132  
Db 97 RCTSRGRRRPFITASGAFAVIAVFLIGYAADLGHLSGSLDKSPKTRAIAPVVGFWIL 156  
QY 133 DFCQGVCTFPLEALLSDLF-RDPDHCROQYSVYAFMISLGGCLGYLLPAID-----W 183  
Db 157 DVANNMLOGPCRALLDLGSQKTRTANALPSPFMAVGNVLGYAAGAVTHLYKLPFT 216  
QY 184 DTSALAPYLTQBECLFGLTLFLTCVAATLLVAEEAALGPTPEAGL-----SAP 235  
Db 217 KTTACDVYCANLKSCTFFIIVLLSLTIVLALSIVYKEK----PWSPPQAVDNAEDDTASQA 272  
QY 236 SLSPHCCPCRARL--AFRNLGALLPRHLQCCMRPTLRRLFVAELCSMMALMTFTLYT 293  
Db 273 SSSAQPMPFFGELGAFKNL-----KEPMWILLVLTCLNIARFPFFLFT 318  
QY 294 DFWGEGLYGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFLVMDRLVQRFTRA 353  
Db 319 DWMGREVYGG--DSSGSABQLKLYDRGVAGALGLMLNSVLTGFTSLGVEVLARGVG-GV 375

## RESULT 10

T14339

sucrose-proton transport protein - carrot

N:Alternate names: sucrose/H+ symporter protein

C:Species: Daucus carota (carrot)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jul-2000

C:Accession: T14339

R:Shakya, R.; Sturm, A.

plant Physiol. 118, 1473-1480, 1998

A:Title: Characterization of source- and sink-specific sucrose/H+ symporters from carrot

A:Reference number: Z17991; MUID:99063785; PMID:9847123

A:Accession: T14339

A:Molecule type: mRNA

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Residues: 1-501 &lt;SHA&gt;

A:Cross-references: EMBL:Y16766; NID:G2969886; PIDN:CAA76367.1; PID:G2969887

A:Experimental source: cultivar Namtaise: leaf

C:Genetics:

A:Gene: SUT1a

C:Superfamily: common tobacco sucrose transport protein

Query Match 10.7%; Score 305.5; DB 2; Length 501;

Best Local Similarity 21.6%; Pred. No. 1.5e-15;

Matches 124; Conservative 91; Mismatches 213; Indels 145; Gaps 14;

QY 18 LLLVNLTLFGLSVCLAAGITYVPPILLLEVGVEEKEFTWVLGIGPVGLVCVPLLGASDH 77

Db 30 LURVASVACGIOFGWALQSLTTPVQELGIPHANSSIWLCPGLSGLLVQPIVGHMSDQ 89

QY 78 WRGYYRRRPFITWALSGLTILSLFLIPRAGWLAGLL--CPDRPLELALLILGVGLLDFC 135

Db 90 CTSKYRRRPFIVAGGTAIIIAHSAIDIGLLGDTADNKTMAIVAFVIGFWILDVA 149

QY 136 GOVCFTFPLEALLSDLF-RDPDHCROQYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLT 194

Db 150 NNMTQCPCRALLADLTGNDARTRVANAYFSLFMAIGNVLGY-----ATGAYSGW 199

QY 195 QBECLFGLTLFLTC-----VAATLLVAEEAL-----GPTPEAGL 232

Db 200 YKVFPSFTSCTTNCANLKSIFYDIITITVISAAKERPRISSQDGPQFSEDGT 259

QY 233 SAPSLSPHCCPCRARLARFRNLGALLPRHLQCCMRPTLRRLFVAELCSMMALMTFTLY 292

Db 260 AQ---SGHI-----EEAFWLWELFGTFRLLPGSVVWVILLVLTCLNWIWNFFPILFD 305

QY 293 TDFVGEGLYGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFLVMDRLVQRFTR 352

Db 306 TDWNGREIYGGEP-----NQQSYSQDGVNMGAFGLMMNSVVLGITSVLMEKLCRTWGS 359

QY 353 AYVLASVAAPFAAGATCLSHSVAVVTASALGTFTFSALQILPYTLASLYHREKQVFLP 412

Db 360 FW-----GLSNILMTTCFFAMLLITFIANKMDY----- 388

QY 413 KYRGDTGGASSEDLSMTSFLPGPKGAPFNGHVAGSGGLPPPPALCGASACDVSVRV 472

Db 389 -----GTNPPN-----GIVISALIVEA 406  
 QY 473 VVGEPTAEV-VP-----GRGICLDLAIDSAFLLSQVAPSLFMGSIQV-----516  
 Db 407 ILGIPLAITYVPYALVSTRISLGLGQSLMGVNLAIWVQVIVSLGSGWDQLFGGG 466  
 QY 517 -SQSVTAYMVSAAGLGLVAIYFATQVVDKSDL 548  
 Db 467 NSPAFVVAALSAPAGLIALIAIRRRPRVDKSL 499

RESULT 11  
 S48788  
 sucrose transport protein - tomato (fragment)  
 C:Species: Lycopersicon esculentum (tomato)  
 C:Date: 27-Jan-1995 #sequence\_revision 27-Jan-1995 #text\_change 09-Jun-2000  
 R:Buerkle, X.Y.Z.; Frommer, W.B.  
 submitted to the EMBL Data Library, October 1994  
 A:Reference number: S48787  
 A:Accession: S48788  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-428 <RUE>  
 A:Cross-references: EMBL:X82275  
 C:Superfamily: common tobacco sucrose transport protein

Query Match 10.4%; Score 297.5; DB 2; Length 428;  
 Best Local Similarity 25.8%; Pred. No. 5.1e-15;  
 Matches 114; Conservative 82; Mismatches 163; Indels 83; Gaps 18;  
 QY 4 RLWVRLRRKRAQLLNLTLFGLVCLAAAGTTPVPLLEVGVEKEWMTVLGIPVL 63  
 Db 25 KLV-----KIVVASIAGVQFGWALQSLTTPVQLLIGIPHRFASFIWLCGPIS 74

QY 64 GLVCVPLLAGSADHWRGRRPFTWALSGLILLSLFLIPRAGWLAGLLCPDP-----R 118  
 Db 75 GMTVQPVGVSDNCSSRRGRRPFTWALSGLILLSLFLIPRAGWLAGLLCPDP-----R 133  
 QY 119 PLELALLIIGVLLDFCGVCFPTLEALLSDFRDPD-HCRQAYSVYAFWISLGGCLGY- 176  
 Db 134 PRAIAVFGVFWLIDVANNMLOGPCRALLDLADLGGKSGKRTANAFPSFKAVGNILGYA 193  
 QY 177 -----LLPAIDWD-TSALAPYLGTQECIF-LGLTLIFLTCVAATLL-----VAEEAALG 224  
 Db 194 AGSYSLFKVFPFSKTKACDMYRANKSCFFIAIFLLSLTTLALTLVRENELPEKEELE 253  
 QY 225 PTEPAEGLSAPSLPHCCPCRLARLAFRNGLALLPRHLQCCMRPTRLRLFVAELCSWMA 284  
 Db 254 IDEKLSGAG-----KSKVPF-FGEIFGALKDL-----PRPMWILLTLVCLNWIA 296

QY 285 LMTFTLFTYDFVGEGLYQGVPRAPGTEARRHYDEGRMGSLGLFLOCAISLVFSLVMDR 344  
 Db 297 WFPFFLYDTDMWAKEVFG-----QVGD--AKLYDLGVRAAGLGLLQSVLVGFMSLGVVF 350  
 QY 345 LVQRFCTRAVYLASVAAPFVAAGATCLSHSAVV-----TASAAATGFT-----388  
 Db 351 LGKKIG-GAKRLWGLNLFVL--AICLAMTILVTKMAEKSGRHDAAGTLMPPTPGVKIGA 406  
 QY 389 ---FSALQI-----LPYTLA 400  
 Db 407 LLLFAALGIPLAVTFSPFALA 428

RESULT 12  
 S38657  
 sucrose transport protein ptpl - common plantain  
 N:Alternate names: sucrose transporter ptpl  
 C:Species: Plantago major (common plantain)  
 C:Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 26-May-2000  
 C:Accession: S38657  
 R:Gahrtz, M.; Stolz, J.; Sauer, N.  
 submitted to the EMBL Data Library, November 1993

A:Reference number: S38657  
 A:Accession: S38657  
 A:Molecule type: mRNA  
 A:Residues: 1-510 <GAH>  
 A:Cross-references: EMBL:X75764; NID:g415987; PIDN:CAA53390.1; PID:g415988  
 C:Genetics:  
 A:Gene: ptpl  
 A:Superfamily: common tobacco sucrose transport protein  
 C:Keywords: sugar transport

Query Match 10.3%; Score 294; DB 2; Length 510;  
 Best Local Similarity 24.7%; Pred. No. 1.1e-14;  
 Matches 126; Conservative 77; Mismatches 199; Indels 108; Gaps 19;  
 QY 18 LLLVNLTLFGLVCLAAAGTTPVPLLEVGVEKEWMTVLGIPVLGVCVPLLAGSADH 77  
 Db 29 IFLVAIAAGVQFGWALQSLTTPVQLLIGIPHKWASIVLWCGPISGMIVQVGVFSDN 88

QY 78 WRGRRRRPFTWALSGLIL-LSLFLIPRAGWLAGL-----LCPDPRLELALLILGV 129  
 Db 89 CTSRFGRRRPF-I-AAGAGLVGVAVVLI---GPAADLGHAGGDSGLGDKLPRAIGVVFVGF 144  
 QY 130 GLLDFCGVCFPTLEALLSDF-RPDHCRQAYSVYAFWISLGGCLGYLLPAID-----182  
 Db 145 WILDVANNMLOGPCRALLDLADLGGKSGKRTANAFPSFKAVGNILGYAAGSYRMYKV 204  
 QY 183 --WDTSAALAPYLGTQEC-LFGLTLIFLTCVAATLL-----VAEEAALGTEPAEGLSA 234  
 Db 205 PRSKTKACDIYCANUKSCFFIISITLITLTLALIVREKRVAREOV---TAAKGGPKI 261

QY 235 PLSLPHCCPCRLARLAFRNGLALLPRHLQCCMRPTRLRLFVAELCSMMALMTFTLYTD 294  
 Db 262 P-----VPELFGALKDLPRPMWILLTLVCLNWIAFWGFLLDFTD 301  
 QY 295 FVGEGLYQGVPRAPGTEARRHYDEGRMGSLGLFLOCAISLVFSLVMDRIVQRF- 350  
 Db 302 WNGREYV-GETOOHKAPELAVIYNKGVSAAGLGLMNSIVLGFASLGVQVMARALGGVKR 360  
 QY 351 -----TRAVLA-SVAAPFVAAGATCLSHSAVVVTSAAALTG--FTFSALQI-----394  
 Db 361 LMGVNFILAIQCLMTIIVTKVASHRPNYSNGVLTQTPSSVKIGALVVFESALGIPLAITF 420  
 QY 395 -LPYTLASLYHREKQVFLPKYRGDTGGASSEDLSL-----MTSFLPGPKG-----438  
 Db 421 SVFPAIASIY-----STTTGSGQSLGLVNLAIVIPQMIVSVASGPMWDMFEG 468  
 QY 439 ----APFNGHVAGGSG-----LLPPPPA 459  
 Db 469 GGNLPAFVVGAVAAASGIFAFTMLPSPPA 498

RESULT 13  
 S51114  
 sucrose-proton symporter - beet  
 C:Species: Beta vulgaris (beet)  
 C:Date: 07-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 26-May-2000  
 C:Accession: S51114  
 R:Westram, A.; Eckhardt, U.; Frommer, W.B.; Riesmeier, J.W.  
 submitted to the EMBL Data Library, January 1995  
 A:Description: Sequence of a sugar beet sucrose transporter cDNA.  
 A:Reference number: S51114  
 A:Accession: S51114  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-523 <WES>  
 A:Cross-references: EMBL:X83850; NID:g633171; PID:g633172  
 C:Superfamily: common tobacco sucrose transport protein

Query Match 10.3%; Score 294; DB 2; Length 523;  
 Best Local Similarity 24.3%; Pred. No. 1.2e-14;  
 Matches 125; Conservative 78; Mismatches 199; Indels 112; Gaps 20;  
 QY 17 QLLVNLTLFGLVCLAAAGTTPVPLLEVGVEKEWMTVLGIPVLGVCVPLLAGSAD 76

Db 38 KLALVASIAAGVQFGWALQSLTPYVQLLGIPTHTWAPIWLCGPISGMIVQPTVGYSD 97  
 Qy 77 HWRGRRRRPFIWALSIGLLSLFLPRAGWLA--GLLCPD-----PRLELALLIGV 129  
 Db 98 RCTSKFRRRPFPI---AVGATLVGFVSLVGLFAADI GHATGDPNGNVKPKRAIAVFGVF 154  
 Qy 130 GLLDPCGQVCFPLEALLSDLFDPD-HCRQAVSVYAFMISLGGCLGY-----LLP 179  
 Db 155 WILDVANNTLQGPCRALADWAGSQAKTRYANAFSFFPMALGNIGYAGSYGLYTVFP 214  
 Qy 180 AIDWTSALAPYLGTQBECLFGLTLFLTCVAATLVABEALGPTPEAGLSAPSLS 239  
 Db 215 FT--HTRACDITYCANLSCCFISITLLIVLTILALSVRER---PFTLDEIQEENLKN 268  
 Qy 240 HCCPCRLARFNLGALLPRHLQCCMRPRLRLFAELCSNMALMTFTLTFDFVGE 299  
 Db 269 NTGGC-ARLPF--FGQFGALKDL-----PKPMLILLVTCNLINIAWPFLLFDMDMGE 321  
 Qy 300 LYQGVPRAEPTGEARRHYDEGRMSGLGLFLQCAISLVSLVMDRLVQRFG--TRAVILA 357  
 Db 322 VYGT-----VGEKAYDNGVHAGALGLMINSVVLGIMSLGIEKLARLVGGVRLWGI 375  
 Qy 358 SVAAPFVAAGATCLSHSVAVVTASA-----ALTF 387  
 Db 376 NLIL-----AVCLAMTI-LVTKSAEHYRATHVPVGAIGPPLPPPGVKGGAIFAFLGI 428  
 Qy 388 TFSALQILPYTLASLYHREK-----QVFLPKYRCDTGGASSESLMT 429  
 Db 429 PLAITTSIPFALASIFSASSGSGQLSLGVNLINIAVVPQMFVSTSGPW-----DAL-- 480  
 Qy 430 SFLPGPKPGAPFNGHVGAGSG-----LLPPPP 458  
 Db 481 -FGGNLPA--FVVGAVATASAILSFLLPPPP 511

RESULT 14

F96741  
 Probable sucrose transporter protein p17M19.4 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
 C/Accession: F96741  
 R/Hirose, T.; Inaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.  
 Plant Cell Physiol. 38, 1389-1396, 1997  
 A/Title: cDNA cloning and tissue specific expression of a gene for sucrose transporter f  
 A/Reference number: T41809; MUID:98182940; PMID:9522469  
 A/Accession: T02982  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-537 <HR>  
 A/Cross-references: EMBL:D87819; NID:G2723470; PIDN:BAA24071.1; PID:G2723471  
 A/Experimental source: cultivar Nipponbare, leaf  
 C/Genetics:  
 A/Gene: SUT1  
 C/Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292.5; DB 2; Length 512;  
 Best Local Similarity 23.8%; Pred. No. 1.5e-14;  
 Matches 120; Conservative 87; Mismatches 211; Indels 87; Gaps 18;  
 Qy 17 QLLVNLTLFGLVCLAGITVPPILLLEGVVEKEFTMTVLGIGVILGVLCVPLIGSASD 76  
 Db 33 KLIISVASIAAGVQFGWALQSLTPYVQLLGIPTHTWAPIWLCGPISGMIVQPTVGYHSD 92  
 Qy 77 HWRGRRRRPFIWALSIGLLSLFLPRAGWLA-----LCPDPRLELALLIGV 129

Db 93 RCBSEFRRRPFIAGVALVAVSVFLI---GFAADMHGSFQDKLENKVRTRAILIIFLTGF 149  
 Qy 130 GLLDPCGQVCFPLEALLSDLFDPD-HCRQAVSVYAFMISLGGCLGY-----LL 178  
 Db 150 WILDVANNTLQGPCRALADWAGSQAKTRYANAFSFFPMALGNIGYAGSYGLYTVFP 209  
 Qy 179 PAIDWTSALAPYLGTQBECLFGLTLFLTCVAATLVABEALGPTPEAGLSAPSLS 238  
 Db 210 PFT--MTKACDIYCANLCTCFPLSITLLIVLTILALSVRER---QWSPQDKEEKS 263  
 Qy 239 PHCCPCRLARFNLGALLPRHLQCCMRPRLRLFAELCSNMALMTFTLTFDFVGE 298  
 Db 264 -----SLFF--FGEIFCAVR---HMKRPMVMLLIIVTINWIAWPFLLYDMDMGR 309  
 Qy 299 GLYQGVPRAEPTGEARRHYDEGRMSGLGLFLQCAISLVSLVMDRLVQRFGTRAVYLA 358  
 Db 310 EVTGG--NSDGRSRSKLYDQGVQAGALGMFNSILLGPFVSLGVESIGRMKG-GAKRLWG 366  
 Qy 359 VAAFPVAAGATCLSHSVAVVTASA-----SAALTGFTFSALQIL----- 395  
 Db 367 CVNFILAI--LMTVLTVTKSAEHRETAGPLAGPSSGIGKAGVFSLFTVLGILAITYS 423  
 Qy 396 -PYTLASLYHREKQVFLPKYRCDTGGASSESLMTSFLFPGKP-----GAPFN---GHVG 447  
 Db 424 IPFALASIFSTNSGAGQLSLGVNLINIAICIPQIMIVSFSSGPLDAQFGGGLNLSFVVGA 483  
 Qy 448 AGSGGL-----PPP--PALCGA 463  
 Db 484 AAVSGVLALTLPSPPPDAPAMSGA 508

RESULT 15

T02982  
 Probable sucrose transporter protein - rice  
 C/Species: Oryza sativa (rice)  
 C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
 C/Accession: T02982  
 R/Hirose, T.; Inaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.  
 Plant Cell Physiol. 38, 1389-1396, 1997  
 A/Title: cDNA cloning and tissue specific expression of a gene for sucrose transporter f  
 A/Reference number: T41809; MUID:98182940; PMID:9522469  
 A/Accession: T02982  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-537 <HR>  
 A/Cross-references: EMBL:D87819; NID:G2723470; PIDN:BAA24071.1; PID:G2723471  
 A/Experimental source: cultivar Nipponbare, leaf  
 C/Genetics:  
 A/Gene: SUT1  
 C/Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292; DB 2; Length 537;  
 Best Local Similarity 26.7%; Pred. No. 1.7e-14;  
 Matches 115; Conservative 72; Mismatches 167; Indels 76; Gaps 19;  
 Qy 17 QLLVNLTLFGLVCLAGITVPPILLLEGVVEKEFTMTVLGIGVILGVLCVPLIGSASD 76  
 Db 51 RLILSGMVAGGVQYQWALQSLTPYVQLLGIPTHTWAPIWLCGPISGMIVQPTVGYHSD 110  
 Qy 77 HWRGRRRRPFIWALSIGLLSLFLI---PRAGWLAIGLCPD-----PRLELALLIL 127  
 Db 111 RCTSKWRRRRPYLTGCVLIGLVVIGFSDIGADYAGMTKEDCSVYHGRWHAAIVVL 170  
 Qy 128 GVGLDFCGQVCFPLEALLSDLF--RDPDHCRQAVSVYAFMISLGGCLGLYLLPAID--- 182  
 Db 171 GPWLLDFSNNTVQGPALMADLSGRHGP--TANSIFCSMMAMGNILYSSGSTNNWH 227  
 Qy 183 -WDTSAALPYLTQBECLFGLTLFLTCVAATLVABE-----AALGPT- 226  
 Db 228 KW-----FPFLKTRACCEACANLKGAFVAVIFLSCLIVITLIFAKEVDFPKGNAAL-PTK 281  
 Qy 227 --EPAGELSAFLSHCCPCRLARFNLGALLPRHLQCCMRPRLRLFAELCSNMA 284

Db	282	SNEPAREGTG-----PLAVLKGRNLTGMPV-----LIVTGL-TWLS	320
Qy	285	LMTFTLFYTDVYCEGLYQGVPR-AEPGTEARRHYDEGRWMSIGLFLQCAISIVFSLVMD	343
Db	321	WPPFILDYTDWNGREIYHGDPKGTDPOIEA---FNOGVRAgapGLLNSIVIGFSSFLIE	377
Qy	344	RLVQREGTRAVYLASVAAPFVAAGATCL-----SHSVA--VVTASAALTGFTFSALQ	393
Db	378	PMCRKVGPRVWVTSNFLVCIAAATAALISFWSLKDFHGTGTVOKAITADKSIKAVCLVLEA	437
Qy	394	ILPYTLASLY	403
Db	438	FLGVPLAVLY	447

Search completed: December 3, 2003, 17:21:14  
Job time : 22 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:16:16 ; Search time 18 Seconds  
(without alignments)  
1444.765 Million cell updates/sec

Title: US-09-593-793a-113

Perfect score: 2861

Sequence: 1 MVQRLVSRLLRHRKAQLK.....AIYFATQVVFVSKSLAKYSA 553

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	489.5	17.1	530	1 MATP_MOUSE	P58355 mus musculus
2	488.5	17.1	530	1 MATP_HUMAN	Q9umx9 homo sapien
3	324.5	11.3	525	1 STP_SPTOL	O03411 spinacia ol
4	235.5	8.2	553	1 SUTI_SCHPO	O14091 schizosacch
5	142	5.0	544	1 YD74_SYNY3	P74168 synecocyst
6	136	4.8	541	1 GT10_HUMAN	O95528 homo sapien
7	125.5	4.4	399	1 TCRI_ECOLI	P02982 escherichia
8	121.5	4.2	503	1 PUR8_STRLP	P42670 streptomyce
9	119.5	4.2	491	1 AMPG_ECOLI	P36670 escherichia
10	117	4.1	440	1 YHJE_ECOLI	P37643 escherichia
11	117	4.1	495	1 SMVA_SALTY	P37594 salmonella
12	111.5	3.9	640	1 Y051_MYCTU	Q10880 mycobacteri
13	111	3.9	473	1 PHDK_NOSCK	O24723 nocardioides
14	110.5	3.9	680	1 CALA_HUMAN	O03692 homo sapien
15	110	3.8	368	1 GALT_HUMAN	O60755 homo sapien
16	110	3.8	606	1 NOOC_THETH	O56227 thermus the
17	109.5	3.8	654	1 SPH2_HUMAN	Q9nra0 homo sapien
18	109	3.8	477	1 YPUM_RHOCA	P26176 rhodobacter
19	107	3.7	419	1 CMLA_PSEAE	P32482 pseudomonas
20	105	3.7	448	1 YJ94_YEAST	P47159 saccharomyc
21	104	3.6	476	1 MELE_SALTY	P30878 salmonella
22	103	3.6	465	1 FTSW_MYCLE	O50186 mycobacteri
23	102.5	3.6	461	1 PUCC_RHOCA	P23462 rhodobacter
24	102	3.6	438	1 SHIA_ECOLI	P76350 escherichia
25	101.5	3.5	428	1 YXIO_BACSU	P42306 bacillus su
26	101.5	3.5	510	1 NANT_YERPE	Q8zch3 yersinia pe
27	101	3.5	437	1 BRAZ_PSEAE	P25185 pseudomonas
28	101	3.5	471	1 MELE_ENTAE	O07366 enterobacte
29	101	3.5	471	1 MELE_KLEPN	Q02581 klebsiella
30	100	3.5	404	1 Y4XM_RHISN	P55705 rhizobium s
31	100	3.5	473	1 Y1HO_SALTY	Q91174 salmonella
32	99	3.5	386	1 CV03_HUMAN	Q9y3p4 homo sapien
33	99	3.5	481	1 LMRA_STRLN	P46104 streptomyce

#### RESULT 1

MATP\_MOUSE  
ID MATP\_MOUSE STANDARD; PRT; 530 AA.

AC P58355;  
DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma)

DE antigen AIM1 (Underwhite protein).

GN MATP OR AIM1 OR UW

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=dy; TISSUE=Eye, Kidney, and Uterus;

RX MEDLINE=21372467; PubMed=11479596;

RA Fukumachi S., Shimada A., Shima A.;

RT "Mutations in the gene encoding B, a novel transporter protein, reduce

RT melanin content in medaka."

RL Nat. Genet. 28:381-385(2001).

RN [2]

RP SEQUENCE FROM N.A., AND VARIANTS UW-DBR ASN-153 AND PRO-435..

RX MEDLINE=21473748; PubMed=11574907;

RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davissom M.T.,

RA King R.A., Brilliant M.H.;

RT "Mutations in the human orthologue of the mouse underwhite gene (uw)

RT underlie a new form of oculocutaneous albinism, OCA4."

RL Am. J. Hum. Genet. 69:981-988(2001).

CC -!- FUNCTION: Melanocyte differentiation antigen. May transport

CC substances required for melanin biosynthesis (By similarity).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By

CC similarity).

CC -!- TISSUE SPECIFICITY: Melanocytes, eyes, kidney and uterus.

CC -!- DISEASE: Defects in MATP are the cause of the UW-dbr phenotype

CC that results in loss of nearly all pigmentation in the homozygous

CC state.

CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AF360357; AAK81713.1; ..

CC MGD; MGI:2153040; Matp.

CC Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;

CC Disease mutation; Albinism.

CC DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).

CC DOMAIN 46 66 1 (POTENTIAL).

CC DOMAIN 67 68 EXTRACELLULAR (POTENTIAL).

CC DOMAIN 69 89 2 (POTENTIAL).

#### ALIGNMENTS

P33607 escherichia  
O87656 salmonella  
O79451 sorex ciner  
P28568 gallus gall  
Q9evn4 pseudomonas  
P79393 bos taurus  
O34245 wolfinella s  
Q02846 homo sapien  
O79452 sorex hayde  
O79969 sorex monti  
O80019 sorex vagra  
O88626 rattus norv

34 98.5 3.4 613 1 NUOL\_ECOLI  
35 98.5 3.4 685 1 FHUB\_SALTY  
36 98 3.4 379 1 CYB\_SORCI  
37 98 3.4 496 1 GTR3\_CHICK  
38 97.5 3.4 354 1 RNFD\_PSEST  
39 97.5 3.4 385 1 P12R\_BOVIN  
40 97.5 3.4 433 1 DCUA\_WOLSU  
41 97.5 3.4 1103 1 CYGD\_HUMAN  
42 97 3.4 336 1 CYB\_SORHA  
43 97 3.4 336 1 CYB\_SORMO  
44 97 3.4 336 1 CYB\_SORVA  
45 97 3.4 370 1 GALT\_RAT

```
FT DOMAIN          90 105 CYTOPLASMIC (POTENTIAL).
FT TRANSEM        106 126 3 (POTENTIAL).
FT TRANSEM        127 138 EXTRACELLULAR (POTENTIAL).
FT TRANSEM        139 159 4 (POTENTIAL).
FT TRANSEM        160 184 CYTOPLASMIC (POTENTIAL).
FT TRANSEM        185 205 5 (POTENTIAL).
FT TRANSEM        206 216 EXTRACELLULAR (POTENTIAL).
FT TRANSEM        217 237 6 (POTENTIAL).
FT TRANSEM        238 318 CYTOPLASMIC (POTENTIAL).
FT TRANSEM        319 339 7 (POTENTIAL).
FT TRANSEM        340 366 EXTRACELLULAR (POTENTIAL).
FT TRANSEM        367 387 8 (POTENTIAL).
FT TRANSEM        388 398 CYTOPLASMIC (POTENTIAL).
FT TRANSEM        399 419 9 (POTENTIAL).
FT TRANSEM        420 425 EXTRACELLULAR (POTENTIAL).
FT TRANSEM        426 446 10 (POTENTIAL).
FT TRANSEM        447 477 CYTOPLASMIC (POTENTIAL).
FT TRANSEM        478 498 11 (POTENTIAL).
FT TRANSEM        499 504 EXTRACELLULAR (POTENTIAL).
FT TRANSEM        505 525 12 (POTENTIAL).
FT DOMAIN         526 530 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD       526 530 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT        153 153 D -> N (IN UN-DBR).
FT VARIANT        435 435 S -> P (IN UN-DBR).
SQ SEQUENCE       530 AA; 57961 MW; 146407916D9FC CRC64;

Query Match
Best Local Similarity 26.1%; Pred. No. 2,7e-28;
Matches 146; Conservative 85; Mismatches 228; Indels 101; Gaps 12;

17 QLLVLLTFLGVLCLAAAGTYVPPLELVGVEKEFTWVLGIPVLGVCVPLGSGASD 76
34 RLVMHSMFMGREGFCYAVEAAYVTPVLLSVGLPKSLYVMVLLSPILFLLQVVGSGASD 93
77 HWRGRRPFTWALSLGILLSLFTPRAGWLAGLLCPDPR-...LELALITLGVGLLD 133
94 HCEARGRRPYTLTALMMLGVALYNGDAVVSALVANPRQKLIWASITWGVVLPD 153
134 FCQGVCTPLLEALLDLDFDPDHCROAYSVYAFMISLGGCLGYLLPAIDWDTALAPYLG 193
154 FSADFDTGPIKAYLFDVCSHQDK-EGKHLVHALFTGGALGYILGAIDWVHLDLGRLG 212
194 TQECFLGLLTFLFCVAATLVAEEAL--GPTEP-----ASGLSAPSLPHCCPC 244
213 TEFQVFFSALVLLICFTLHLCISIPEAPLRDAATDPPSQDPPQSGSLSGASGHYEY---- 268
245 RARLAFRNLGA-----LLPRLHQLCCRMPTRLRLFLVABLCSW 282
269 GSIEKVKNGGADTEQPVQEWKMKKPSQSQRTWSMKSLLRALVNMPSHYRCLCVSHLIGW 328
283 MALMTFTLFTYDFVGBGLYOGVPRAEPGTERRHYDEGVPMGSLGLFLQCAISLVFLVM 342
329 TAFLSNMLFTDFMGQIVVHGDPYGAHNSTEFILYERGVGVCWGLCINSVFSSVSYFQ 388
343 DRLVQRFGRTRAVLASVAAPPAAGATCLSHSVAVVTASAAALGFTFSALQILPYTLASL 402
389 KAMVSYGLKGLYFMGVLPLGLGTGFIGLFPNYSITLVLCSMFVMSSTLYVFPFLIAE 448
403 YHREKVFLPKYRGDTGGASSEDLSMTSFLPLGPKGAPFPNGHVAGGSGLLPPPPALCG 462
449 YHREEE-----KEKGQEA-----PGPDNQGR-----DCA 477
463 ASACDVSVRVVGEPTEARVVPGRGICLDLAILDSAPLLSQVAPSLFMGSIVOLSSQVTA 522
478 ALTQVQL-----AQILVGG-----LGLFLNMAGSVVV 506
523 YNVSAGLGLVAIYEQTVV 542
507 VVITASAVSLIGCCFVALFV 526
```

RESULT 2  
MATP\_HUMAN

```
ID AC Q9UMX9; Q9BTM3; MATP_HUMAN STANDARD; PRT; 530 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma
DE antigen AIM1).
DE MATP OR AIM1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A. (ISOFORM AIM-1A).
RC TISSUE=Melanoma;
RX MEDLINE=21115844; PubMed=11221837;
RA Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.;
RT "Use of an in vitro immunoselected tumor line to identify shared
RT melanoma antigens recognized by HLA-A*0201-restricted T cells.";
RL Cancer Res. 61:1089-1094 (2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM AIM-1B).
RC TISSUE=Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ussid T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Richards S., Worley K.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [3]
RP ALTERNATIVE SPLICING.
RA Ferro S.;
RN Unpublished observations (NOV-2001).
RN [4]
RP DISEASE AND VARIANT LEU-374.
RX MEDLINE=21473748; PubMed=11574907;
RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davison M.T.,
RA King R.A., Brillant M.H.;
RT "Mutations in the human orthologue of the mouse underwhite gene (uw)
RT underlie a new form of oculocutaneous albinism, OCA4.";
RL Am. J. Hum. Genet. 69:981-988 (2001).
CC -!- FUNCTION: Melanocyte differentiation antigen. May transport
CC substances required for melanin biosynthesis (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
CC similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=AIM-1a;
CC IsoId=Q9UMX9-1; Sequence=Displayed;
CC Name=AIM-1b;
CC IsoId=Q9UMX9-2; Sequence=VSP_006297, VSP_006298, VSP_006299;
CC Name=AIM-1c;
CC IsoId=Q9UMX9-3; Sequence=VSP_006296;
CC -!- TISSUE SPECIFICITY: Expressed in most melanoma cell lines and
CC melanocytes.
CC -!- DISEASE: Defects in MATP are the cause of oculocutaneous albinism
CC type 4 (OCA4). OCA4 is characterized by hypopigmentation of skin,
```





```
DR Pfam: PF00083; sugar_tr; 1.  
DR TIGRfam: TIGR01301; GPH_sucrose; 1.  
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE NEG.  
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE NEG.  
KW Transmembrane; Transport; Sugar transporter; Symport.  
FT DOMAIN 1 37 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 38 58 1 (POTENTIAL).  
FT TRANSMEM 72 92 2 (POTENTIAL).  
FT TRANSMEM 107 127 3 (POTENTIAL).  
FT TRANSMEM 145 165 4 (POTENTIAL).  
FT TRANSMEM 184 204 5 (POTENTIAL).  
FT TRANSMEM 230 250 6 (POTENTIAL).  
FT TRANSMEM 295 315 7 (POTENTIAL).  
FT TRANSMEM 338 358 8 (POTENTIAL).  
FT TRANSMEM 373 393 9 (POTENTIAL).  
FT TRANSMEM 422 442 10 (POTENTIAL).  
FT TRANSMEM 455 475 11 (POTENTIAL).  
FT TRANSMEM 488 508 12 (POTENTIAL).  
FT DOMAIN 509 525 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 525 AA; 54992 MW; 018347A4D2C1C66 CRC64;  
  
Query Match 11.3%; Score 324.5; DB 1; Length 525;  
Best Local Similarity 23.2%; Pred. No. 2.5e-16;  
Matches 129; Conservative 91; Mismatches 196; Indels 141; Gaps 20;  
  
QY 23 LITFGLVCLAAGITY-----VPLLLEGVVEKEFMVWLGIPVGLVLCVPLLSGA 74  
DB 35 LKKLGLVASVAGVQFGWALQLSLTPYVQLLGPHTWAAVYIWLCPISGMIVQPLVGY 94  
QY 75 SDHWRGVRGRRPFIWALSIGLLSLFLIPRAGWLA--GLLCDDP-----RPLELALLIL 127  
DB 95 SDRCTSRFGRRRPFIAGAAALVAVAVGLI---GFAADIGAASGDPGNVAKPRAIYV 151  
QY 128 GVGLLDFCGVCVPTPEALSLDLR-DPDHCRQAVSYAFMISLGLGCLGYLLPAID- 182  
DB 152 GFWILDVANNTLOGPCRALLDADWAGSQTTRYANAPFFSFMALGNIGVAAGSYSRLYT 211  
QY 183 -----WDTSAIAPYLGTOEBCFLGILT-LIPLTCAVATLLVAEEAALGPTPEAGLSAPSL 237  
DB 212 VFPFTKTAACDVYCANLKSCFFISITLLIIVLTILSLSVKVERQITIDEIQEEDDLKRN 271  
QY 238 SPHCCPCRLAFRNIGALLPRHLQCCRRPTRLRLFVAELCSWMLMTFTLFYTDVFG 297  
DB 272 SSCG-----ARLPF--FGQLGALKDL-----PKWMLILLVLTALNWTANWFFLLPDTDW 321  
QY 298 EGYLQGVPRAPGPTGARHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFQ--TRAVY 355  
DB 322 KEVYGGT-----VGEKLYDQGVHAGALGMINSVVLGVMSLSIEGLARMVGGAKRLWG 375  
QY 356 LASVAAPPVAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415  
DB 376 IVNIIL-----AVCLAMTV-LVTKSAE-----HFRDSHHIM----- 405  
QY 416 GDTGGASSEDLSMTSFLPGPKGAPFNGHVAGGSGLLPPPALCGASCDVSVRVVWG 475  
DB 406 -----GSAVPPPPPA--GVKGGALAIFAVLG 429  
QY 476 EPTEARV-VP-----GRGICLDLAILDSAFLLSQV-----AP 506  
DB 430 IPLAITFSIPFALASIFSSSGSGQLSLGLVNLAIIVVQMFVSVTSGPMDAMFGGGLNP 489  
QY 507 SLFMGSIVQLSQSVTAY 523  
DB 490 AFVVGAVAATASAVLSF 506  
  
RESULT 4  
ID SUT1 SCHPO STANDARD; PRT; 553 AA.  
AC O14031; Q9UU60;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)
```

General alpha-glucoside permease.  
SUT1 OR SPAC2F3.08.  
Schizosaccharomyces pombe (Fission yeast).  
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
Schizosaccharomycetales; Schizosaccharomycetaceae;  
Schizosaccharomycetes.  
NCBI\_TaxID=4896;  
[1]  
SEQUENCE FROM N.A.  
RP STRAIN=972;  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,  
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,  
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,  
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Fumelle B.,  
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,  
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
RT "The genome sequence of Schizosaccharomyces pombe";  
RL Nature 415:871-880(2002).  
[2]  
SEQUENCE OF 35-166 FROM N.A.  
RP STRAIN=968 h90;  
RX MEDLINE=20223868; PubMed=10759889;  
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,  
RA Hirakawa Y.;  
RT "Large-scale screening of intracellular protein localization in living  
fission yeast cells by the use of a GFP-fusion genomic DNA library";  
RL Genes Cells 5:169-190(2000).  
[3]  
IDENTIFICATION, AND FUNCTION.  
RX MEDLINE=21065106; PubMed=11136464;  
RA Reinders A., Ward J.M.;  
RT "Functional characterization of the alpha-glucoside transporter Sut1p  
from Schizosaccharomyces pombe, the first fungal homologue of plant  
sucrose transporters";  
RL Mol. Microbiol. 39:445-454(2001).  
CC -!- FUNCTION: Responsible for the transport of maltose and sucrose  
into the cell, with the concomitant export of a proton (Symport  
system).  
CC -!- PATHWAY: Maltose and sucrose metabolism.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Z99165; CAB16264.1; -.  
DR EMBL; AB027797; BAA87101.1; -.  
DR PIR; T38541; T38541.  
DR GenedB.SPombe; SPAC2F3.08; -.  
KW Transmembrane; Transport; Sugar transporter; Symport.





Db 109 GFAISLSMCCIYVSELV-GPQRQGVLSYELVAGITVIGLLSYALNYALAGTWPWGRHM 167  
QY 187 ---ALAPYLGTQBECLFGLLTFLTCVAATLLVAEEALGPTEPREGLSAPLSHCCP 243  
Db 168 FGWATAP-----AVLQSLSLFLP--AGT---DETATKDLIPQGGAPKLGGF--- 212  
QY 244 CRARLAFRNIGALLPRHLQJCCMPRTLRFLVAELCSNMALMTFTLFYDFVGEGLYQG 303  
Db 213 -RPYSFGLDLFRARDNMRG---RTTVGLGLVLQFQLTGQPNVLCYA--STIFSVGFHGG 266  
QY 304 VPRAEPTTEARRHYDEGVRCMSGLFLQCAISLVFSLVMDRLVQRCTRAVYLASVA--- 360  
Db 267 -----SSAVLASVGL---GAVKVAATLTAMGLVDGRALLAGCALMA 308  
QY 361 -----APFVAAGATCLSHSVAVVTASAALTG---FTFSALQILPYTLASLYHR 405  
Db 309 LSVSGIGLVSPFVPMDSGPSCL--AVPNATGQGLGDSGLLODLSLPIPT-----N 360  
QY 406 EKQVFLPKYRGDTGGASSEDLSMTSFLPGPKPGAPFPNGHVAGSGGLLPPPPALCGASA 465  
Db 361 EDQ-----REPILSTAKTKPHPRSGDPSAPPRALSSALPGPP----- 399  
QY 466 CDVSVRVVVEPTTEARVPPRG-----ICLDLALDSAF-----LLSQVAP-- 506  
Db 400 -----LPARGHALLRWTTALLCLMVFAFSFGFGFTWLVLSIYPVE 442  
QY 507 -----SLFMG-----STVQLSQSVTAVMVSA--GLGLVAVY 536  
Db 443 IGRAPAFCSFNWAAANLFLSFLDLIGTIGLSWTFLLYGLTAVLGLGYILF 496

## RESULT 7

TCR1\_ECOLI : STANDARD; PRT; 399 AA.  
AC P02982; :  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Tetracycline resistance protein, class A (TET(A)).  
GN TET.  
OS Escherichia coli.  
OG Plasmid RPI.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TRANSPOSON=Tn1721;  
RX MEDLINE=92192465; PubMed=1312499;  
RA Allmeier H., Cresnar B., Greck M., Schmitt R.;  
RT "Complete nucleotide sequence of Tn1721: gene organization and a  
RT novel gene product with features of a chemotaxis protein.";  
RL Gene 111:11-20 (1992).  
RN [2]  
RP SEQUENCE OF 85-399 FROM N.A.  
RC STRAIN=DH1.  
RA Fletcher J.N., Hart C.A., Batt R.M., Saunders J.R.;  
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC PLASMID=RP1; TRANSPOSON=Tn1721;  
RX MEDLINE=83299270; PubMed=6310527;  
RA Waters S.H., Rogowsky P., Grinstead J., Altenbuchner J., Schmitt R.;  
RT "The tetracycline resistance determinants of RPI and Tn1721:  
RT nucleotide sequence analysis.";  
RL Nucleic Acids Res. 11:6089-6105 (1983).  
RN [4]  
RP TOPOLOGY.  
RX MEDLINE=92388137; PubMed=1517220;  
RA Allard J.D., Bertrand K.P.;  
RT "Membrane topology of the pBR322 tetracycline resistance protein.  
RT TetA-PhoA gene fusions and implications for the mechanism of TetA  
RT membrane insertion.";

Query Match 4.4%; Score 125.5; DB 1; Length 399;  
Best Local Similarity 22.3%; Pred. No. 0.051;  
Matches 104; Conservative 57; Mismatches 136; Indels 169; Gaps 27;  
QY 19 LVLNLLTFGLVCLAAAGITYVPPL---LLEVGVEKEFMT---LLEVGVEKEFMT---MWLGIPVLGLVCPVLL 71  
Db 7 LVLILSTVALD---AVGIGLIMPVLPGLLRLDVHSNDVTAHVIGILLALYALMQFACAPVL 63

J. Biol. Chem. 267:17809-17819 (1992).  
-!- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE  
EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE  
ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN  
FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTI-PORTER.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X61367; CA43643.1; -  
DR EMBL; L29404; AA83545.1; -  
DR EMBL; X00006; CA24909.1; -  
DR PIR; A03509; YTECR1.  
DR PIR; JQ1479; JQ1479.  
DR InterPro; IPR004734; Drug\_resist.  
DR InterPro; IPR007114; MFS.  
DR InterPro; IPR005828; Sub\_transporter.  
DR InterPro; IPR005829; Sug\_transporter.  
DR InterPro; IPR001958; TCR\_TetA.  
DR Pfam; PF00083; sugar\_tr; 1.  
DR PRINTS; PRO1035; TCR\_TETRA.  
DR TIGRFAMs; TIGR00880; 2\_A\_01\_02; 1.  
DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; UNKNOWN 1.  
KW Antibiotic resistance; Transmembrane; Inner membrane; Transport;  
KW Antiport; Ion transport; Hydrogen ion transport; Transposable element;  
KW Plasmid.  
FT DOMAIN 1 7 CYTOPLASMIC (PROBABLE).  
FT TRANSMEM 8 27 PROBABLE.  
FT DOMAIN 28 45 PERIPLASMIC (PROBABLE).  
FT TRANSMEM 46 66 PROBABLE.  
FT DOMAIN 67 79 CYTOPLASMIC (PROBABLE).  
FT TRANSMEM 80 100 PROBABLE.  
FT DOMAIN 101 103 PERIPLASMIC (PROBABLE).  
FT TRANSMEM 104 124 PROBABLE.  
FT DOMAIN 125 138 CYTOPLASMIC (PROBABLE).  
FT TRANSMEM 139 159 PROBABLE.  
FT DOMAIN 160 160 PERIPLASMIC (PROBABLE).  
FT TRANSMEM 161 181 PROBABLE.  
FT DOMAIN 182 210 CYTOPLASMIC (PROBABLE).  
FT TRANSMEM 211 231 PROBABLE.  
FT DOMAIN 232 246 PERIPLASMIC (PROBABLE).  
FT TRANSMEM 247 267 PROBABLE.  
FT DOMAIN 268 277 CYTOPLASMIC (PROBABLE).  
FT TRANSMEM 278 298 PROBABLE.  
FT DOMAIN 299 320 PERIPLASMIC (PROBABLE).  
FT TRANSMEM 300 320 PROBABLE.  
FT DOMAIN 321 339 CYTOPLASMIC (PROBABLE).  
FT TRANSMEM 340 360 PROBABLE.  
FT DOMAIN 361 364 PERIPLASMIC (PROBABLE).  
FT TRANSMEM 365 385 PROBABLE.  
FT DOMAIN 386 399 CYTOPLASMIC (PROBABLE).  
FT CONFLICT 5 5 R -> I (IN REF. 3).  
FT CONFLICT 55 55 M -> V (IN REF. 3).  
FT CONFLICT 75 75 V -> I (IN REF. 3).  
FT CONFLICT 84 84 A -> T (IN REF. 3).  
FT CONFLICT 201 203 ASF -> SFV (IN REF. 3).  
SQ SEQUENCE 399 AA; 42240 MW; 298427E6B5478374 CRC64;

QY 72 GSASDHWGRYGR-----RPEIWSLGILLSLFLIPRAGWLAGL 112  
Db 64 GALS-----RFGRRPVLVSLAGAAVDYAIMATAPFLWLYIGRIVA----- 106  
QY 113 LCPDPRLELALLILGVLLDGCQVCFPTLEALLSDLFDPDHCRCQAVSVVAFMISLGG 172  
Db 107 -----GITGATGAVA-----GAYIADITDGERARH-----GCFM-----SA 138  
QY 173 CLGYLLPAIDMTSALAPYLGQEBCLFGLLTFLITCVAAITLLVAEEAALGPTPEAGL 232  
Db 139 CFFGFGMVA-----GPVLG-----GLM----- 154  
QY 233 SAPSLPHCCPCARLAFRNLG-----ALLPRHLQCCMRPTLRLFVAELCS--W--- 282  
Db 155 --GGFSPH-APFFAAAALNGLNFLTGCFLLPESHK---GERRPLRREALNPLASFWRARG 208  
QY 283 ---MALMTFTFLFYDTPVGBGLYQGVRA--EPGTEARRHYDE---GVRMGSLGLFLQCAI 334  
Db 209 MTVVAALMAVFFIMQLVGQ-----VPAALWVIFGEDRFHWDATTIGISLAAGILHSLAQ 263  
QY 335 SLVFSLVMDRLVQRFOTRAVYLASVA---APFVAAGAT---CLSHSVAVVTASAAULTGFTF 389  
Db 264 AMITGPVAARLGER---RALMLGMIADGTGYILLAPATRGWMAFFIMVLLASG---GIGM 317  
QY 390 SALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSLPGP 435  
Db 318 PALQAM---LSROVBERQ-----GOLQGSIAULTSLTSIV-GP 352  
  
RESULT 8  
PUB8\_STRLP STANDARD; PRT; 503 AA.  
AC P42670;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Puromycin resistance protein pub8.  
GN PUB8.  
OS Streptomyces lipmanii (Streptomyces alboniger).  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomyceinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=132472;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 12461;  
RX MEDLINE=94109397; PubMed=7916693;  
RA Tercero J.A., Lacalle R.A., Jimenez A.;  
RT "The pub8 gene from the pur cluster of Streptomyces alboniger encodes  
a highly hydrophobic polypeptide which confers resistance to  
puromycin.";  
RL Eur. J. Biochem. 218:963-971 (1993).  
CC -1- FUNCTION: MAY BE INVOLVED IN ACTIVE PUROMYCIN EFFLUX ENERGIZED BY  
A PROTON-DEPENDENT ELECTROCHEMICAL GRADIENT. IN ADDITION, IT COULD  
BE IMPLICATED IN SECRETING N-ACETYLPUROMYCIN. THE LAST  
INTERMEDIATE OF THE PUROMYCIN BIOSYNTHESIS PATHWAY, TO THE  
ENVIRONMENT.  
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
AS THE DRUG RESISTANCE TRANSLUCASE FAMILY).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X76855; CAA54186.1; --  
DR InterPro; IPR004638; Efflux\_EmrB.  
DR InterPro; IPR007114; MFS.  
DR InterPro; IPR005828; Sub\_transporter.

DR InterPro; IPR001411; TCR\_TetB.  
DR Pfam; PF00083; sugar\_tx; 1.  
DR PRINTS; PR01036; TCR\_TetB.  
DR TIGRFAMs; TIGR00711; efflux\_EmrB; 1.  
KW Antibiotic resistance; Antibiotic biosynthesis; Transmembrane;  
Transport.  
FT DOMAIN 1 24 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 25 45 POTENTIAL.  
FT DOMAIN 46 64 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 65 85 POTENTIAL.  
FT DOMAIN 86 92 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 93 113 POTENTIAL.  
FT DOMAIN 114 122 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 123 143 POTENTIAL.  
FT DOMAIN 144 152 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 153 173 POTENTIAL.  
FT DOMAIN 174 181 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 182 202 POTENTIAL.  
FT DOMAIN 203 212 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 213 233 POTENTIAL.  
FT DOMAIN 234 241 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 242 262 POTENTIAL.  
FT DOMAIN 263 280 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 281 301 POTENTIAL.  
FT DOMAIN 302 313 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 314 334 POTENTIAL.  
FT DOMAIN 335 346 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 347 367 POTENTIAL.  
FT DOMAIN 368 371 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 372 392 POTENTIAL.  
FT DOMAIN 393 422 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 423 443 POTENTIAL.  
FT DOMAIN 444 461 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 462 482 POTENTIAL.  
FT DOMAIN 483 503 CYTOPLASMIC (POTENTIAL).  
SQ SEQUENCE 503 AA; 51852 MW; FA3B814DF9A3EB74 CRC64;  
  
Query Match 4.2%; Score 121.5; DB 1; Length 503;  
Best Local Similarity 23.2%; Pred. No. 0.13;  
Matches 131; Conservative 61; Mismatches 175; Indels 197; Gaps 32;  
  
QY 26 FGLVECLAAGITYV-----PPLLEVGVESEKFTMTV-----LGTPVLGLVCVPL 70  
Db 26 WGLVILAAQLLVLDGTVNIALPSVQDRLGMSDTSROWVITAYTLAFGLL-----L 79  
QY 71 LGSASDHWGR-----YGRRRPFIWALSIGILLSLFLIPRAGW-LAGLL---CPDPRLEL 122  
Db 80 LG-----GRVADAFGRRIE-----AVGIL-----GFLASLLGGAAPDPTLFL 119  
QY 123 ALLILGVLLDGCQVCFPTLEALLSDLFDPDHCRCQAVSVVAFMISLGGCLGYLLPAID 182  
Db 120 ARALQV-----FAALAPAL-ALINTLFTPEGERGKAFGVVGVGGGAUVLLAG--- 171  
QY 183 WDTALAPYLGQEBCLFGLLT--LIFITCVAAITLLVAEEAALGPTPEAGLSAPLSPH 240  
Db 172 -----GLLTYLDMRWCLVNPVALLLIG----- 197  
QY 241 CPCRARLAFRNLGALLPRHLQCCMRPTLRLFVAEL--CSWMAIMFTFLFYDTPVGE 298  
Db 198 --CR-----LLPRDR-----RTGRAVLDPGLTLLGCGGLVAIVYAF----- 232  
QY 299 GLYQGVRAEPGTEARRHYDEG--VRMGSGLFLQCAISLVFSLVMDRLVQRFOTRAVYL 356  
Db 233 -----AEESGWDPLVLLVGLVLMVAFALVRRVQDPLP----- 271  
QY 357 ASVAAPFPVAAGATCLSHSVAVVTASAAULTG-----FTFSALQILPYT--LASLYHREKQV 410  
Db 272 PGVVAHRVGGG-----FLVVGLPQIGLFLFLTYLYLQGLIDYSPVLTV-----AF 320  
QY 411 LPKYRGDTGGASSEDLSMTSLFPGKP-----GAPPNGHVGGSGGL-----PPPALC 461  
Db 321 LPLGLGIAGVSS-----LIAARLLPRTLRPTLIVGALL-----AAAGMALTRLEPDP 370



```
Db 156 MLVSGGLAWL--ADKW-----LGWQ--MYWLMAALLIPCIATILAPE-----PT 198
Qy 227 EPAEGLSAPSLPHCCPCRLAFRLNIGALLRLHLQCCMRPTLRRLRFAVEL-----279
Db 199 D-----TIP-----VPKLEQAVVAFRPFGRN 222
Qy 280 CSWMLMTFTL-----FYTFVGEGLYQGVPAEFGTEARRHYDGVGMGLFLQCA 333
Db 223 NAWLILLVLYKGDFAFMSLTTLIRGV-----GFDAG---EVGVNKTGLLATIV 274
Qy 334 ISLVSLVNDRLVQRGTGTRAVLYASVAAPFAAGATCLSHVAVVTASALTGFTFSALQ 393
Db 275 GALYGILMQLSL-----FRALLIFI-----LQASNAGYW 307
Qy 394 ILPYTLASLYHREKQVFLPKYRGDTGCGASEDSMTSLFLPGPKGAPFNGHVAGGSGI 453
Db 308 LLSITDKHLYSMGAUVFENLCCGMGTSAFVALLMT-----LCKNKSFSATQPAL 356
Qy 454 LPPPPALCGASACDVSRVVVVGPEPTARVVVPGRGICLDLAILDSAPLLSQVA---PSLFM 510
Db 357 L-----SALSAVGRVVG-PVAGWVFAHGW-----STYLSVAAPVGLIL 398
Qy 511 GSIVQLSQ-----SVTAYMVS-----AAGLGLVAIYFATQVVFQKSLAKYS 552
Db 399 LLVCROTLEVTRVNDNFISRTAYPAGYAFAMWTLAAGVSLIAWLLL-LTMDALDLTHFS 457

RESULT 10
YHJE_ECOLI
ID YHJE_ECOLI STANDARD; PRT; 440 AA.
AC P37643;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical metabolite transport protein yhjE.
GN YHJE OR B3523.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(potential).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
H.INFLUENZAE HI0281 AND HI0418.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U00039; AAB18499.1; -
CC EMBL; AS000429; AAC76548.1; -
CC PIR; S47743; S47743.
CC EcoGene; EGI2249; yhjE.
CC InterPro; IPR004736; Cit_H_symport.
CC InterPro; IPR007114; MFS.
CC InterPro; IPR005828; Sub_transporter.
CC InterPro; IPR005829; Sug_transporter.
CC Pfam; PF00083; sugar_1.
CC TIGRfam; TIGR00883; 2A0106; 1.
CC PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
CC PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.

KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
Complete proteome.
KW DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 35 55 1 (POTENTIAL).
FT DOMAIN 56 66 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 67 87 2 (POTENTIAL).
FT DOMAIN 88 108 3 (POTENTIAL).
FT TRANSMEM 109 129 4 (POTENTIAL).
FT TRANSMEM 130 150 5 (POTENTIAL).
FT TRANSMEM 151 167 6 (POTENTIAL).
FT TRANSMEM 168 188 7 (POTENTIAL).
FT TRANSMEM 189 192 8 (POTENTIAL).
FT TRANSMEM 193 213 9 (POTENTIAL).
FT TRANSMEM 214 248 10 (POTENTIAL).
FT TRANSMEM 249 269 11 (POTENTIAL).
FT TRANSMEM 270 289 12 (POTENTIAL).
FT TRANSMEM 290 310 13 (POTENTIAL).
FT TRANSMEM 311 320 14 (POTENTIAL).
FT TRANSMEM 321 341 15 (POTENTIAL).
FT TRANSMEM 342 345 16 (POTENTIAL).
FT TRANSMEM 346 366 17 (POTENTIAL).
FT TRANSMEM 367 384 18 (POTENTIAL).
FT TRANSMEM 385 405 19 (POTENTIAL).
FT TRANSMEM 406 410 20 (POTENTIAL).
FT TRANSMEM 411 431 21 (POTENTIAL).
FT TRANSMEM 432 440 22 (POTENTIAL).
SQ SEQUENCE 440 AA; 47208 MW; 3A23302A9DCB5BE CRC64;

Query Match 4.1%; Score 117; DB 1; Length 440;
Best Local Similarity 28.4%; Pred. No. 0.24;
Matches 48; Conservative 29; Mismatches 54; Indels 38; Gaps 10;

Qy 53 MTMLVIGIGVPLGVCVPLLGASDHWGRGRRRPPFWALSGLILLSLFLI-PRAGWLAG 111
Db 295 MWAVIGFG-----VMPVAGLLAD-----AFGRKSKSVITILLFALFAPNPLG----341
Qy 112 LLCPPDRP-LELALLILGVLDLFCQVCFTPLEALLSLDFRDPDHCROAYSVAFMISL 170
Db 342 ----SGNPILVPAFLLLGLSLMG-----LTFGPMGALLPELF--PTEVR--YTGASPSYNV 389
Qy 171 GCGIGYLLPAIDWDTSAAPYLGTOECLFGLLTL-IFLTCVAATILVA 218
Db 390 ASILG-----ASVAPYIAAWLQNTYNGILGAVGLYLAANAGLTLLA 428

RESULT 11
SMVA_SALTY
ID SMVA_SALTY STANDARD; PRT; 495 AA.
AC P37594;
DT 01-OCT-1994 (Rel. 30, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Methyl viologen resistance protein smva.
GN SMVA OR STM1574.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SL1303;
RX MEDLINE=95011654; PubMed=7926834;
RA Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA Ichimura S., Noda Y.;
RT "The methyl viologen-resistance-encoding gene smva of Salmonella
typhimurium."
RL Gene 148:173-174(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.B., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
```



RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
LT2"; 413:852-856(2001).  
RL Nature 413:852-856(2001).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
CC (Potential).  
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D26057; BAA05055.1; -.  
DR EMBL; AE008769; AAL20492.1; -.  
DR StvGene; SG10384; smva.  
DR InterPro; IPR007114; MFS.  
DR InterPro; IPR001411; TCR-TetB.  
DR PRINTS; PR01036; TCR-TetB.  
KW Transmembrane; Inner membrane; Complete proteome.  
FT TRANSMEM 5 25 POTENTIAL.  
FT TRANSMEM 44 64 POTENTIAL.  
FT TRANSMEM 73 93 POTENTIAL.  
FT TRANSMEM 96 116 POTENTIAL.  
FT TRANSMEM 135 155 POTENTIAL.  
FT TRANSMEM 158 178 POTENTIAL.  
FT TRANSMEM 192 212 POTENTIAL.  
FT TRANSMEM 220 240 POTENTIAL.  
FT TRANSMEM 260 280 POTENTIAL.  
FT TRANSMEM 299 319 POTENTIAL.  
FT TRANSMEM 327 347 POTENTIAL.  
FT TRANSMEM 357 377 POTENTIAL.  
FT TRANSMEM 391 411 POTENTIAL.  
FT TRANSMEM 459 489 POTENTIAL.  
FT CONFLICT 144 144 A -> R (IN REF. 1).  
FT CONFLICT 147 147 G -> V (IN REF. 1).  
FT CONFLICT 182 182 V -> D (IN REF. 1).  
FT CONFLICT 198 198 A -> V (IN REF. 1).  
FT CONFLICT 227 227 F -> Y (IN REF. 1).  
FT CONFLICT 447 447 L -> Y (IN REF. 1).  
FT CONFLICT 454 457 AILD -> GKLT (IN REF. 1).  
SQ SEQUENCE 495 AA; 52134 MW; 961F77C748CDA164 CRC64;

Query Match  
Best Local Similarity 4.1%; Score 117; DB 1; Length 495;  
Matches 92; Conservative 60; Mismatches 151; Indels 136; Gaps 22;

QY 17 QLLVNLVTFGL-----EVCLAAGITVYVPLLLEVG-----EKFMT 54  
DB 72 RLILMGTLFGLASLAASHTASWLIATRVLLAIGAAIMVPAVL-AGIRATPCBEKHN 130  
QY 55 MVLGI-----GPVLG-----IVCVPL-----IGSADHWGRYGR- 85  
DB 131 MALGVAAVVGSGAAGFGLTGILLSHFYWGSVFLNVPVLMGLTARYVPRQAGRD 190  
QY 86 RPF-----IWALS-----GILLSLFLIPRAGWLAGL 112  
DB 191 QPLNLGHAVNLIITAILLYVSATKALGHLSLWISFTLLTGALLGLFIRQLATSRPM 250  
QY 113 LCDPRLPLEALILVGLLDPCGVCFTPLEALLSDFRDPDCHQAYSVYAFMI----- 168  
DB 251 I--DMELFTHRIILSGV-VNMAWTAMITLVGFELMAQELQV-HGLSPYEAGVFMVPMV 306  
QY 169 -----SLGGCL-----GYLLPAIDWDTSAAPYLGTOECLFGLLTFLTC 210  
DB 307 ASFGSGPIAGVLVSRGLRLVATGGMALSALSPYGLAMTDF-STOQWQAWGLMALIGFS- 364

QY 211 VVATLLVBEAALGPTPEAE-GLSAPSLSPHCCPCRARLAFRNGLGALLRLHOLCCMRP 269  
DB 365 --AASALLASTSAIMAAAPAEKAAAGAIETMAYELGAGLGIAIFGLLSRFSASIRLPA 423  
QY 270 TLRRLFVACLCSWMLMTFTLFTYDFVGG--LYQGVP--RAEPTGTAEARRH---YDEGVR 322  
DB 424 GLEAQSIARASSM-----GEAVQLANSPLPTQQAIIIDAAARHAFIWSHVA 470  
QY 323 MGSGLG-LFLQCAISLVFSL 340  
DB 471 LSSAGSMILLAVGWMFSL 489

RESULT 12  
Y051 MYCTU  
ID Y051 MYCTU STANDARD; PRT; 640 AA.  
AC Q10880; OS3628;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical protein Rv0083.  
GN Rv0083 OR MT0090 OR MTCV251.01 OR MTV030.27.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia P.,  
RA Badcock K., Baaham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of *Mycobacterium tuberculosis* from the  
RT complete genome sequence.";  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of *Mycobacterium tuberculosis* clinical and  
RT laboratory strains.";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE  
CC OF CHLOROPLASTS OR MITOCHONDRIA.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL021428; CAAL6264.1; -.  
DR EMBL; AE006920; AAK44315.1; -.  
DR PIR; D70850; D70850.  
DR TIGR; MT0090; -.  
DR Tuberculist; Rv0083; -.  
DR InterPro; IPR003918; NADH\_oxred.  
DR InterPro; IPR001750; Oxidored q1.  
DR Pfam; PF00361; oxidored q1; 1.  
DR PRINTS; PR01437; NUOXDRDTASE4.

KW Hypothetical protein; Oxidoreductase; Transmembrane;  
 KW Complete proteome.  
 FT TRANSMEM 8 28 POTENTIAL.  
 FT TRANSMEM 52 72 POTENTIAL.  
 FT TRANSMEM 90 110 POTENTIAL.  
 FT TRANSMEM 136 156 POTENTIAL.  
 FT TRANSMEM 179 199 POTENTIAL.  
 FT TRANSMEM 208 228 POTENTIAL.  
 FT TRANSMEM 241 261 POTENTIAL.  
 FT TRANSMEM 277 297 POTENTIAL.  
 FT TRANSMEM 298 318 POTENTIAL.  
 FT TRANSMEM 352 372 POTENTIAL.  
 FT TRANSMEM 391 411 POTENTIAL.  
 FT TRANSMEM 446 466 POTENTIAL.  
 FT TRANSMEM 497 517 POTENTIAL.  
 FT TRANSMEM 619 639 POTENTIAL.  
 SQ SEQUENCE 640 AA; 65627 MW; 6254C289DBD108A8 CRC64;  
 Query Match 3.9%; Score 111.5; DB 1; Length 640;  
 Best Local Similarity 22.3%; Pred. No. 0.87;  
 Matches 129; Conservative 73; Mismatches 222; Indels 155; Gaps 29;  
 QY 16 AQLLV-----NLTFLGL--EVCLAGITYV-----PPLLEVGVEKFTMTVLGIGPVLG 64  
 Db AAMLVPAAGSVTTFLAWELMAISLILVLSHARPQVRSAGLWYAVMTQLGFIALVIG 152  
 QY 65 LVCVPLGASDHWGRYGRRRPFIWALSIGLILSLF-----LIPRAGWLAGL 112  
 Db LVCVPLGASDHWGRYGRRRPFIWALSIGLILSLF-----LIPRAGWLAGL 112  
 QY 153 LVVLAAG--GSDRFAG-----LGAVCDGVRAAVFMTLVGFGSKAGLVLPHAWLPRA 203  
 Db LVVLAAG--GSDRFAG-----LGAVCDGVRAAVFMTLVGFGSKAGLVLPHAWLPRA 203  
 QY 113 LCPDPRP-----LEALLILGV--GLLDFCGQVCTPLEALLSDFRDPDHCROAYSVYAFM 167  
 Db HPEAPSPVSAALMAAMVNGIYGVFDLQL-----LCPGRWGLA-----L 246  
 QY 168 ISLIG---CLGYLLPAIDMTSALAPYLGTQBECLFGLTLPLTCVAATLVAEBAALG 224  
 Db LVCVPLGASDHWGRYGRRRPFIWALSIGLILSLF-----LIPRAGWLAGL 112  
 QY 247 LAVGGSYALYGVLAQSVAAADLKELAYSTEN---MGLITL-----ALGAATLFADTGAYG 299  
 Db LAVGGSYALYGVLAQSVAAADLKELAYSTEN---MGLITL-----ALGAATLFADTGAYG 299  
 QY 225 PTEPAGLSAPLSPPCCPCRLAPRNLGALLPR-----LHOLCCRMPTLRLFLV 276  
 Db PASIAAAAAAMLHMAHA--AFKSLAFMAAGSVLAATGLRDLGLGLARRMPAT----- 351  
 QY 277 AELCSMWALMTPLFY-----TDFVGEGLYGVVPRAAPGTEARRHYDEGVR 322  
 Db LAVGGSYALYGVLAQSVAAADLKELAYSTEN---MGLITL-----ALGAATLFADTGAYG 299  
 QY 352 -----TVFVGVAALGACGLPLGAGFVSEMLVQSLIHAAPGHPDPIVALTPLA 399  
 Db LAVGGSYALYGVLAQSVAAADLKELAYSTEN---MGLITL-----ALGAATLFADTGAYG 299  
 QY 323 MGSGLFLQCAISLVFS-----LVMDRLVQRFGRTRAVYLASVAAFVFAAGATCLSHSVA 376  
 Db VGVVALATGLSVAAMTKATFGIGFLARPRSTQAEAREAPASMRAGMAIAAGA-CLVLAVA 458  
 QY 377 --VVTASALTGTFPSALQILPVT--LASLYHREKQVFLPKYRGDTGGASDSMTSFLP 433  
 Db VGVVALATGLSVAAMTKATFGIGFLARPRSTQAEAREAPASMRAGMAIAAGA-CLVLAVA 458  
 QY 459 PLLVAPMVRRAAATLPAQAQVKTGLGAV-----VRLPAMSG-----SIAP 499  
 Db VGVVALATGLSVAAMTKATFGIGFLARPRSTQAEAREAPASMRAGMAIAAGA-CLVLAVA 458  
 QY 434 GPRKPAFPNGHVGAGSGGLP-----PPPA-----LCGASACDVSRVVVVGPEPTAR 481  
 Db VGVVALATGLSVAAMTKATFGIGFLARPRSTQAEAREAPASMRAGMAIAAGA-CLVLAVA 458  
 QY 500 GVTAADV-----AAALAVAVLARWRFRPPAPARLPFLWACGA--DLTVWQYATISFAE 553  
 Db VGVVALATGLSVAAMTKATFGIGFLARPRSTQAEAREAPASMRAGMAIAAGA-CLVLAVA 458  
 QY 482 VVFGRCIGLDLALDSAFLLSQVAPSLFMGSIQVLSQSV 520  
 Db PLQVFGDVLRPDPTDIEVTHTAESRYMAERTYRTAV 590  
 RESULT 13  
 PHDK NOCSK  
 ID PHDK NOCSK STANDARD; PRT; 473 AA.  
 AC O24723;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable 1-hydroxy-2-naphthoate transporter.  
 OS Nocardioides sp. (strain KP7).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Propionibacterineae; Nocardioidaceae; Nocardioides.  
 OX NCBI\_TaxID=35761;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97474276; PubMed=9335300;  
 RA Iwabuchi T., Harayama S.;  
 RT "Biochemical and genetic characterization of 2-carboxybenzaldehyde  
 dehydrogenase, an enzyme involved in phenanthrene degradation by  
 Nocardioides sp. strain KP7";  
 RL J. Bacteriol. 179:6488-6494 (1997).  
 CC -!- FUNCTION: PROBABLY INVOLVED IN THE TRANSPORT OF 1-HYDROXY-2-  
 NAPHTHOATE.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane  
 (Potential).  
 CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AB000735; BAA23264.1; -;  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR005828; Sub\_transporter.  
 DR InterPro; IPR005829; Sug\_transporter.  
 DR Pfam; PF00083; sugar\_tr; 1.  
 DR PROSITE; PS00216; SUGAR\_TRANSPORT\_1; 1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; FALSE\_NEG.  
 KW Transport; Transmembrane; Inner membrane.  
 FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 26 46  
 FT DOMAIN 47 59 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 60 80 2 (POTENTIAL).  
 FT DOMAIN 81 91 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 92 112 3 (POTENTIAL).  
 FT DOMAIN 113 121 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 122 142 4 (POTENTIAL).  
 FT DOMAIN 143 153 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 154 174 5 (POTENTIAL).  
 FT DOMAIN 175 180 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 181 201 6 (POTENTIAL).  
 FT DOMAIN 202 233 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 234 284 7 (POTENTIAL).  
 FT DOMAIN 285 304 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 305 325 8 (POTENTIAL).  
 FT DOMAIN 326 330 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 331 351 9 (POTENTIAL).  
 FT DOMAIN 352 354 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 355 375 10 (POTENTIAL).  
 FT DOMAIN 376 398 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 399 418 11 (POTENTIAL).  
 FT DOMAIN 419 421 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 422 444 12 (POTENTIAL).  
 FT DOMAIN 445 473 CYTOPLASMIC (POTENTIAL).  
 SQ SEQUENCE 473 AA; 49109 MW; D6D765D376260D8A CRC64;  
 Query Match 3.9%; Score 111; DB 1; Length 473;  
 Best Local Similarity 24.3%; Pred. No. 0.69;  
 Matches 114; Conservative 66; Mismatches 180; Indels 110; Gaps 27;  
 QY 12 RHKQALLVNLTF--GLEVCLAGITYVPLLELVGVEKFTMTVLGIGVLCVCP 69  
 Db RHKQALLVNLTF--GLEVCLAGITYVPLLELVGVEKFTMTVLGIGVLCVCP 69  
 QY 22 RRQRIIVLMVFLIMVADGMDITLASHL--FPFVIRDWGVPSAVTLVSLGVVMAIGAL 79  
 Db RRQRIIVLMVFLIMVADGMDITLASHL--FPFVIRDWGVPSAVTLVSLGVVMAIGAL 79  
 QY 70 LLGSASDHWGRYGRRRPFIWALSIGLILSLFLIPRAG--WLAGLLCPDPRLEALLILG 128  
 Db LLGSASDHWGRYGRRRPFIWALSIGLILSLFLIPRAG--WLAGLLCPDPRLEALLILG 128  
 QY 80 VSGPVDNRW--GRKG-----VTVGVFV--LFLATAGLGLTG----DIHSPALRIISC 125  
 Db VSGPVDNRW--GRKG-----VTVGVFV--LFLATAGLGLTG----DIHSPALRIISC 125  
 QY 129 VGLLDFCGQVCTPLEALLSDFRDPDHCRC--QAYSVYAFMISLGGCLGYLLPAIDWDTSA 187  
 Db VGLLDFCGQVCTPLEALLSDFRDPDHCRC--QAYSVYAFMISLGGCLGYLLPAIDWDTSA 187

Db 126 FGL-----GAVMPVAL-TIVADM--PKARRAQMVSIAPAGVGVGSIIGAYLAA-----A 172

Qy 188 LAPVLGTQEBCLF-GLTLTFLFTCAATLLVABEA-----ALGPTBPAEG 231

Db 173 VIPTLQWQVVLATLAPLILPFFVA--LVPEPAIISVVRGIPPEARIRSAALVALVAPDRD 230

Qy 232 LSAPSLSPHCCPCARAFNAGLIPRLHOLCCMRPTRLRLRFLVABLCSWMAAMTFTTLF 291

Db 231 IAGVDLT-----RAGL-----TLGAGEVR-----AKALFAELICRPLLGVTLLIW 270

Qy 292 YTFPVGEG-----LYQGV-----RAEPCTEARRHYDEGVRMGSLGLFLOCAISLVFSLVMD 343

Db 271 GVFFVQSGLLVLQVPMMLQAPAGLST---VESGLIVAMTGWG-----ALIGQUTTA 332

Qy 344 RLQRFQFTRAVYLASVAAPFAAGATCLSHSVAVVTASAAALTGTF-----SALQI-L 395

Db 323 FILKRF-DRIALAAFIWFSV-----VGLLIVAAFGTGFGRFGYFTLLFAIGLSL 371

Qy 396 PYTLASLYHREKOVFLPKYR-----GDTGGASSDSLSMTSLPQKPGCAPF 441

Db 372 PATAAAMQSVTTLAYBEEPRATGWSAGFAGRLGTLTYGALGOTLLIGAGF 421

RESULT 14

CAIA\_HUMAN

ID CAIA\_HUMAN STANDARD; PRT; 680 AA.

AC Q03632;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Collagen alpha 1(X) chain precursor.

GN COL10A1.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92109659; PubMed=1764025;

RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,

RA Solomon E., Grant M.E., Boot-Handford R.P.;

RT "The human collagen X gene. Complete primary translated sequence and

RT chromosomal localization.";

RL Biochem. J. 280:617-623(1991).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=93012005; PubMed=1397333;

RA Reichenberger E., Beier F., Luvalle P., Olsen B.R., von der Mark K.,

RA Bertling W.M.;

RT "Genomic organization and full-length cDNA sequence of human collagen

RT X.";

RL FEBS Lett. 311:305-310(1992).

RN [3]

RP SEQUENCE FROM N.A.

RA Beier F., Lammi M.B., von der Mark K.;

RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Williams S.;

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE OF 52-680 FROM N.A.

RX MEDLINE=92267014; PubMed=1587271;

RA Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;

RT "Cloning of the human and mouse type X collagen genes' and mapping of

RT the mouse type X collagen gene to chromosome 10.";

RL Eur. J. Biochem. 206:217-224(1992).

RN [6]

RP SEQUENCE OF 561-666 FROM N.A.

RX MEDLINE=91243838; PubMed=2037056;

RA Apte S., Mattei M.-G., Olsen B.R.;

RT "Cloning of human alpha 1(X) collagen DNA and localization of the

RT COL10A1 gene to the q21-q22 region of human chromosome 6.";

RL FEBS Lett. 282:393-396(1991).

RN [7]

RP SEQUENCE OF 547-655 FROM N.A.

RX MEDLINE=92077285; PubMed=1743401;

RA Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;

RT "In situ hybridization studies on the expression of type X collagen

RT in fetal human cartilage.";

RL Dev. Biol. 148:562-572(1991).

RN [8]

RP REVIEW ON VARIANTS.

RX MEDLINE=97255959; PubMed=9101290;

RA Kuivaniemi H., Tromp G., Prockop D.J.;

RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-

RT associated collagen (type IX), and network-forming collagen (type X)

RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";

RL Hum. Mutat. 9:300-315(1997).

RN [9]

RP VARIANTS SMCD ASP-598 AND PRO-614.

RX MEDLINE=94136476; PubMed=8304336;

RA Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,

RA Grant M.E., Boot-Handford R.P.;

RT "Amino acid substitutions of conserved residues in the

RT carboxyl-terminal domain of the alpha 1(X) chain of type X collagen

RT occur in two unrelated families with metaphyseal chondrodysplasia

RT type Schmid.";

RL Am. J. Hum. Genet. 54:169-178(1994).

RN [10]

RP VARIANT SMCD ARG-591.

RX MEDLINE=94272470; PubMed=8004099;

RA McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;

RT "Additional mutations of type X collagen confirm COL10A1 as the

RT Schmid metaphyseal chondrodysplasia locus.";

RL Hum. Mol. Genet. 3:303-307(1994).

RN [11]

RP VARIANT SMCD VAL-618.

RX MEDLINE=95181449; PubMed=7876225;

RA Chan D., Cole W.G., Rogers J.G., Bateman J.F.;

RT "Type X collagen multimer assembly in vitro is prevented by a Gly618

RT to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid

RT metaphyseal chondrodysplasia.";

RL J. Biol. Chem. 270:4558-4562(1995).

RN [12]

RP VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.

RX MEDLINE=95331767; PubMed=7607655;

RA Bonaventure J., Chaminade F., Maroteaux P.;

RT "Mutations in three subdomains of the carboxy-terminal region of

RT collagen type X account for most of the Schmid metaphyseal

RT dysplasias.";

RL Hum. Genet. 96:58-64(1995).

RN [13]

RP VARIANT SMCD PRO-600.

RX MEDLINE=96375754; PubMed=8782043;

RA Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,

RA Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;

RT "Mutations within the gene encoding the alpha 1(X) chain of type X

RT collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but

RT not several other forms of metaphyseal chondrodysplasia.";

RL J. Med. Genet. 33:450-457(1996).

RN [14]

RP VARIANTS SMCD GLU-18 AND ARG-18.

RX MEDLINE=97220591; PubMed=9067753;

RA Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;

RT "Mutations in the N-terminal globular domain of the type X collagen

RT gene (COL10A1) in patients with Schmid metaphyseal

RT chondrodysplasia.";

RL Hum. Mutat. 9:131-135(1997).

RN [15]

RP VARIANTS SMD GLU-595.

RX MEDLINE=99057503; PubMed=9837818;

RA Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,

RA Nakamura Y.;

RT "Mutation of the type X collagen gene 'COL10A1' causes

RT spondylometaphyseal dysplasia.";

Am. J. Hum. Genet. 63:1659-1662 (1998).  
[16]  
RN VARIATION SMCD CYS-597.  
RX MEDLINE=99069781; PubMed=9852679;  
RA Sawai H., Ida A., Nakata Y., Koyama K.;  
RT "Novel missense mutation resulting in the substitution of tyrosine by  
RT cysteine at codon 597 of the type X collagen gene associated with  
RT Schmid metaphyseal chondrodysplasia.";  
RL J. Hum. Genet. 43:259-261 (1998).  
CC -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC  
CC CHONDROCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE  
CC MINERALIZATION ZONES OF HYALINE CARTILAGE.  
CC -!- SUBUNIT: Homotrimer.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
CC unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE  
CC METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED  
CC DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE  
CC PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.  
CC RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE  
CC METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE  
CC KNEES.  
CC -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL  
CC DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE  
CC SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE  
CC VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.  
CC -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.  
CC -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; X60382; CAA42933.1; -;  
DR EMBL; X65120; CAA46236.1; -;  
DR EMBL; X98568; CAA67178.1; -;  
DR EMBL; AL121963; CAB87590.1; -;  
DR EMBL; S68531; AAC60615.1; -;  
DR EMBL; X58879; CAA41686.1; -;  
DR EMBL; W74050; CAA61221.1; -;  
DR EMBL; X72579; CAA51170.1; -;  
DR EMBL; X72580; CAA51170.1; JOINED.  
DR PIR; S26396; CGHUID.  
DR PDB; 1GR3; 14-FEB-02.  
DR Genew; HGNC:2185; COL10A1.  
DR MIM; 120110; -;  
DR MIM; 156500; -;  
DR MIM; 184250; -;  
DR GO; GO:0005581; C:collagen; TAS.  
DR GO; GO:0005202; F:collagen; TAS.  
DR GO; GO:0001501; P:skeletal development; TAS.  
DR InterPro; IPR001073; C1q.  
DR Pfam; PF00386; C1q; 1.  
DR Pfam; PF01391; Collagen; 8.  
DR PRINTS; PR00007; COMPLEMENTC1Q.  
DR SMART; SM00110; C1Q; 1.  
DR PROSITE; PS01113; C1q; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cartilage; Collagen; Signal; Disease mutation; Polymorphism;  
KW 3D-structure.  
FT SIGNAL 1 18 POTENTIAL.  
FT CHAIN 19 680 COLLAGEN ALPHA 1(X) CHAIN.  
FT DOMAIN 19 56 NONHELICAL REGION (NC2).  
FT DOMAIN 57 519 TRIPLE-HELICAL REGION.  
FT DOMAIN 520 680 NONHELICAL REGION (NC1).  
FT DOMAIN 545 680 C1Q.  
FT VARIATION 18 18 G -> E (IN SMCD).  
FT FTId=VAR\_001838.

FT VARIANT 18 18 G -> R (IN SMCD).  
FT FTId=VAR\_001839.  
FT VARIANT 545 545 G -> R.  
FT FTId=VAR\_001840.  
FT VARIANT 591 591 C -> R (IN SMCD).  
FT FTId=VAR\_001841.  
Query Match 3.9%; Score 110.5; DB 1; Length 680;  
Best Local Similarity 30.7%; Pred. No. 1.1; Indels 35; Gaps 7;  
Matches 43; Conservative 11; Mismatches 51;  
QY 412 PKYRGDTG--GASSSDSLMTSFLPGP--KPGAPFNGHVAGGSGLLPPPPALCGASACD 467  
DB 463 FGSKGDPGSPGPPGAGIATKGLNGTGPDPGPPGRHSGEPG---LPGPP----- 510  
QY 468 VSVRVVVGEPTEARVVGPGICGLDIALDSAPLSQVAPSLFMGSIQVLSQVYAMVSA 527  
DB 511 -----GPPGP-----PGQ-----AVMEPGFIKAGORPSLSGTPLVSANOGVTGMPVSA 553  
QY 528 AGLGLVAIY--PATQVVEFK 545  
DB 554 FTVILSKAYPAIGTPIPFDK 573  
RESULT 15  
GALT\_HUMAN  
ID GALT\_HUMAN STANDARD; PRT; 368 AA.  
AC O60755;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DE 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Galanin receptor type 3 (GAL3-R) (GALR3).  
GN GALR3 OR GALNR3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98389766; PubMed=9722565;  
RA Smith K.E., Walker M.W., Artymyshyn R., Bard J., Borowsky B.,  
RA Tamm J.A., Yao W.-J., Vaysse P.J.-J., Brancheck T.A., Gerald C.,  
RA Jones K.A.;  
RT "Cloned human and rat galanin GALR3 receptors: pharmacology and  
RT activation of G-protein inwardly rectifying K+ channels.";  
RL J. Biol. Chem. 273:23321-23326 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Uterus;  
RA Bennett M.M., Lescoe M.K., Gallipoli P.Z., Ramabhadran T.V.;  
RT "Homologue of the human galanin 2 receptor gene isolated from a human  
RT uterus cDNA library.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBSJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99048961; PubMed=9832121;  
RA Kolakowski L.F. Jr., O'Neill G.P., Howard A.D., Broussard S.R.,  
RA Sullivan K.A., Feighner S.D., Sawdzargo M., Nguyen T., Kargman S.,  
RA Shiao L.-L., Hreniuk D.L., Tan C.P., Evans J., Abramovitz M.,  
RA Chateaufneuf A., Coulombe N., Ng G., Johnson M.P., Tharian A.,  
RA Khoshbouei H., George S.R., Smith R.G., O'Dowd B.F.;  
RT "Molecular characterization and expression of cloned human galanin  
RT receptors GALR2 and GALR3.";  
RL J. Neurochem. 71:2239-2251 (1998).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99127120; PubMed=9928159;  
RA Lismaa T.P., Fathi Z., Hort Y.J., Iben L.G., Dutton J.L., Baker E.,  
RA Sutherland G.R., Shine J.;  
RT "Structural organization and chromosomal localization of three human  
RT galanin receptor genes.";  
RL Ann. N.Y. Acad. Sci. 863:56-63 (1998).  
RN [5]

```

SEQUENCE FROM N.A.
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA Clump M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grahm D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA Martyn I.D., Maashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.B., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA Scheet P., Walker C., Wansley A., Wohldmann P., Pepin K., Nelson J.,
RA Korf I., Bedell J.A., Hillier L., Wardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelman L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Seroussi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,
RA Wilkison P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
CC -1- FUNCTION: RECEPTOR FOR THE HORMONE GALANIN.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; AF073799; AAC35944.1; -
DR ENBL; AF067733; AAC18860.1; -
DR ENBL; AF129514; AAD47348.2; -
DR ENBL; AF129513; AAD47348.2; JOINED.
DR ENBL; Z97630; -; NOT ANNOTATED_CDS.
DR Genew; HGNC:4134; GALR3.
DR MIM; 603692; -
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0005886; C:plasma membrane; TAS.
DR GO; GO:0004966; F:galanin receptor activity; TAS.
DR GO; GO:0007631; P:feeding behavior; TAS.
DR GO; GO:0007611; P:learning and/or memory; TAS.
DR GO; GO:0007194; P:negative regulation of adenylate cyclase ac. .; TAS.
DR GO; GO:0007218; P:neuropeptide signaling pathway; TAS.
DR GO; GO:0007268; P:synaptic transmission; TAS.
DR InterPro; IPR000276; GPCR_Rhodopsin.

DR PFam; PF00001; 7tm 1; 1.
DR PRINTS; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_2; 1.
DR PROSITE; PS00232; G_PROTEIN_RECEP_F1_2; 1.
DR KW G-protein coupled receptor; Transmembrane; Glycoprotein;
DR Multi-gene family; Lipoprotein; Palmitate.
DR DOMAIN 1 20 EXTRACELLULAR (POTENTIAL).
DR FT TRANSMEM 21 41 1 (POTENTIAL).
DR FT DOMAIN 42 57 CYTOPLASMIC (POTENTIAL).
DR FT TRANSMEM 58 78 2 (POTENTIAL).
DR FT DOMAIN 79 96 EXTRACELLULAR (POTENTIAL).
DR FT TRANSMEM 97 118 3 (POTENTIAL).
DR FT DOMAIN 119 138 CYTOPLASMIC (POTENTIAL).
DR FT TRANSMEM 139 159 4 (POTENTIAL).
DR FT DOMAIN 160 184 EXTRACELLULAR (POTENTIAL).
DR FT TRANSMEM 185 205 5 (POTENTIAL).
DR FT DOMAIN 206 236 CYTOPLASMIC (POTENTIAL).
DR FT TRANSMEM 237 257 6 (POTENTIAL).
DR FT DOMAIN 258 259 EXTRACELLULAR (POTENTIAL).
DR FT TRANSMEM 260 280 7 (POTENTIAL).
DR FT DOMAIN 281 368 CYTOPLASMIC (POTENTIAL).
DR FT CARBOHYD 6 6 N-LINKED (GLCNAC. .) (POTENTIAL).
DR FT DISULFID 95 172 BY SIMILARITY.
DR FT LIPID 308 308 PALMITATE (BY SIMILARITY).
DR FT DOMAIN 218 225 POLY-ALA.
DR FT DOMAIN 310 318 POLY-ARG.
DR SQ SEQUENCE 368 AA; 39573 MW; E8F46B66422C3150 CRC64;

Query Match 3.8%; Score 110; DB 1; Length 368;
Best Local Similarity 21.9%; Pred. No. 0.63;
Matches 94; Conservative 35; Mismatches 145; Indels 156; Gaps 17;

QY 87 PF1WALSGLILSLFLIPRAG---WLAGLLCPDPRLE-----LALLILGVGLDRCGV 138
DB 21 PVVFAI-----IFLLGTGNGLVAVLLQPCPSAWQPGSTTDLFILNLAVADLCFIL 73
QY 139 CFTPLLEALLSDFRDPDHCRCQAVSVVAFMISLGGCLGYLLPAIDMTSALAPYLGTOEC 198
DB 74 CCVFPQATI-----YTLDA--W----- 88
QY 199 LFGLLT-----LIFLTCTVAATLLVAEEAALGFTPEAGLSAPSLSPHCCPCARLARPN 252
DB 89 LFGALVCKAVHLIIVLTWYASFT-----LAASVDRY-----LAVR- 125
QY 253 LGALLPRHLQCRMPRTLRLFLVLAELCSWMAIMFTLEYTFVGEGLYQGVPRABPGTE 312
DB 126 -----HPLRSRALRTPRNARAAGVGLVLLAALFSAFSLSYGTGYGALCVPawe 177
QY 313 ARHYDEGVRMGSLGLFLOCA-ISLVFSLVMDRLVQRFQ-----TRAVYLASVAA 361
DB 178 DARRALDVATTAAGVLLPVAVVSLAYGTLRFLWAANGVAGAAAARRRATGRAGRAM 237
QY 362 FPVAA-----GATCLSHSVAVVTTAALTGFTTFSALQILPYTLASLY 403
DB 238 LAVALYALCMGPHALILCFWYGRFAFSPATYACRLASHCLAYANSLNPLVYALASRH 297
QY 404 -----HREKQVFLPKYRGDTGCGASSELSMTSFLPGPKPGAPFPNGHV 447
DB 298 FRARFRLWPCGRRRRHARRA-LRRVRPASSG-----PPGCGDARPSGRLL 344
QY 448 AGGSGLLPPP 457
DB 345 AGG-GQGPEP 353

Search completed: December 3, 2003, 17:19:49
Job time : 21 secs

```

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:16:46 ; Search time 40 Seconds  
(without alignments)  
3567.577 Million cell updates/sec

Title: US-09-593-793a-113

Perfect score: 2861

Sequence: 1 MVQRLWVSRLLRHRKAQLL.....AIYFATQVDPKSLAKYSA 553

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phage.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_rvirus.\*  
16: sp\_bacteriap.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2861	100.0	553	Q96JT2	Q96jt2 homo sapien
2	2798	97.8	553	Q95KI5	Q95ki5 macaca fasc
3	2602	90.9	553	Q8XOH7	Q8xoh7 mus musculus
4	2531	88.5	501	Q95KC5	Q95kc5 macaca fasc
5	2099	73.4	450	Q8K252	Q8k252 mus musculus
6	519.5	18.2	599	Q9VSV1	Q9vsv1 drosophila
7	512	17.9	112	Q8RI10	Q8ri10 mus musculus
8	484.5	16.9	530	Q8C204	Q8c204 mus musculus
9	456	15.9	576	Q90Z74	Q90z74 oryzias lat
10	446	15.6	751	Q8K4S3	Q8k4s3 rattus norv
11	394.5	13.8	754	Q8MMT9	Q8mmt9 dictyosteli
12	347.5	12.1	515	Q65803	Q65803 daucus caro
13	347.5	12.1	515	Q9FNR6	Q9fnr6 daucus caro
14	346.5	12.1	501	Q9SQK6	Q9sqk6 vitis vinif
15	346.5	12.1	515	Q9SLN7	Q9sln7 daucus caro
16	346	12.1	612	Q9SQK5	Q9sqk5 vitis vinif

17	338	11.8	512	10	Q9S7Z5	Q9s7z5 apium grave
18	337.5	11.8	523	10	Q04077	Q04077 vicia faba
19	336.5	11.8	501	10	Q9SP63	Q9sp63 vitis vinif
20	333	11.6	512	10	Q9ZTB9	Q9ztb9 apium grave
21	332.5	11.6	508	10	Q8VIX3	Q8vix3 brassica ol
22	330	11.5	516	10	Q43653	Q43653 solanum tub
23	329.5	11.5	524	10	Q9XHL6	Q9xhl6 pisum sativ
24	328.5	11.5	506	10	Q944W2	Q944w2 oryza sativ
25	328.5	11.5	594	10	Q80605	Q80605 arabidopsis
26	327.5	11.4	506	10	Q948L0	Q948l0 oryza sativ
27	325.5	11.4	604	10	Q9FVL6	Q9fv16 lycopersico
28	324.5	11.3	502	10	Q9SP14	Q9sp14 alonsoa mer
29	324.5	11.3	512	10	Q8RWQ6	Q8rwq6 arabidopsis
30	323.5	11.3	512	10	Q39231	Q39231 arabidopsis
31	323.5	11.3	512	10	Q80550	Q80550 arabidopsis
32	317.5	11.1	507	10	Q40583	Q40583 nicotiana t
33	317.5	11.1	530	10	Q9M535	Q9mv35 euphorbia e
34	316.5	11.1	500	10	Q9FVJ6	Q9fvj6 lycopersico
35	313	10.9	521	10	Q9SXM0	Q9sxm0 zea mays (m
36	311	10.9	533	10	Q41152	Q41152 ricinus com
37	309.5	10.8	607	10	Q8LPM4	Q8lpm4 citrus sine
38	307.5	10.7	488	10	Q9FV92	Q9fv92 solanum tub
39	306.5	10.7	513	10	Q8VIX4	Q8vix4 brassica ol
40	306.5	10.7	528	10	Q8LPM6	Q8lpm6 citrus sine
41	306	10.7	523	10	Q9M422	Q9m422 hordeum vul
42	305.5	10.7	501	10	Q65929	Q65929 daucus caro
43	302.5	10.6	510	10	Q9SP15	Q9sp15 asarina bar
44	302	10.6	523	10	Q8RUL3	Q8rul3 triticum ae
45	301.5	10.5	429	10	Q40167	Q40167 lycopersico

## ALIGNMENTS

RESULT 1

Q96JT2 PRELIMINARY; PRT; 553 AA.  
ID Q96JT2  
AC Q96JT2;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE Prostein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Prostate;  
RX MEDLINE=21139094; PubMed=11245466;  
RA Xu J., Kalos M., Stolk J.A., Zasloff E.J., Zhang X., Houghton R.L.,  
RA Filho A.M., Nolasco M., Badaro R., Reed S.G.;  
RT "Identification and characterization of prostein, a novel prostate-  
RT specific protein."  
RL Cancer Res. 61:1563-1568(2001).  
DR EMBL: AV033593; AAK54386.1; --  
SQ SEQUENCE 553 AA; 59322 MW; 0AFA23FBC742A667 CRC64;

Query Match 100.0%; Score 2861; DB 4; Length 553;  
Best Local Similarity 100.0%; Pred. No. 2.6e-209;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLNVLLTFGLVCLAAQITVPPLLLLVGVGVEEKFTWVLGIG 60  
DB 1 MVQRLWVSRLLRHRKAQLLNVLLTFGLVCLAAQITVPPLLLLVGVGVEEKFTWVLGIG 60  
QY 61 PVLGLVCPPLGASDHWGRGRRRPFIWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120  
DB 61 PVLGLVCPPLGASDHWGRGRRRPFIWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120  
QY 121 ELALLILGVLLDFCGQVCFTPLEALLSDLFRDPDHCQAYSVYAFMISLGSCLYLLPA 180  
DB 121 ELALLILGVLLDFCGQVCFTPLEALLSDLFRDPDHCQAYSVYAFMISLGSCLYLLPA 180

QY 181 IDWTSALAPYLGTOBECFLGGLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPLSHP 240  
 DB 181 IDWTSALAPYLGTOBECFLGGLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPLSHP 240  
 QY 241 CCPCARLARAFNLGALLPRHLQCCMPRTLRLFVAELCSWMLMTFTFLFYTDFVGEGL 300  
 DB 241 CCPCARLARAFNLGALLPRHLQCCMPRTLRLFVAELCSWMLMTFTFLFYTDFVGEGL 300  
 QY 301 YQVPAEAPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQVPAEAPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
 DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540  
 QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

## RESULT 2

Q95K15 PRELIMINARY; PRT; 553 AA.  
 ID Q95K15  
 AC Q95K15  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE Hypothetical 59.4 kDa protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Macaca.  
 OX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Temporal cortex;  
 RA Oeada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,  
 RA Suzuki Y., Sugano S., Hashimoto K.;  
 RT "Isolation of full-length cDNA clones from macaque brain cDNA  
 RT libraries";  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB060851; BAB46871.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 553 AA; 59392 MW; 0718F3A91FB3BF1E CRC64;

Query Match 97.8%; Score 2798; DB 6; Length 553;  
 Best Local Similarity 98.2%; Pred. No. 1.6e-204;  
 Matches 542; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 MVQRLVWSRLRRKRAQLLLVLLTFGLVCLAAAGTTPVPLLVGVGVEEKFMTWVLGIG 60  
 DB 1 MVQRLVWSRLRRKRAQLLLVLLTFGLVCLAAAGTTPVPLLVGVGVEEKFMTWVLGIG 60  
 QY 61 PVGLVVCVPLLSASDHWRGRRRPFITWALSGLILLSLFLIPRAGWLAGLCPDPRPL 120  
 DB 61 PVGLVVCVPLLSASDHWRGRRRPFITWALSGLILLSLFLIPRAGWLAGLCPDPRPL 120  
 QY 121 ELALLTLGVLLDFCCQVCFPLEALLSDLPDPDHCRAQYSYAPWISLGGCLGYLLPA 180  
 DB 121 ELALLTLGVLLDFCCQVCFPLEALLSDLPDPDHCRAQYSYAPWISLGGCLGYLLPA 180  
 QY 181 IDWTSALAPYLGTOBECFLGGLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPLSHP 240

DB 181 IDWTSALAPYLGTOBECFLGGLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPLSHP 240  
 QY 241 CCPCARLARAFNLGALLPRHLQCCMPRTLRLFVAELCSWMLMTFTFLFYTDFVGEGL 300  
 DB 241 CCPCARLARAFNLGALLPRHLQCCMPRTLRLFVAELCSWMLMTFTFLFYTDFVGEGL 300  
 QY 301 YQVPAEAPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQVPAEAPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 APVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
 DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIFATQ 540  
 QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

## RESULT 3

Q8K0H7 PRELIMINARY; PRT; 553 AA.  
 ID Q8K0H7  
 AC Q8K0H7  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DE Hypothetical protein (PROSTEIN homolog).  
 GN 2210413P12RIK.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 DR EMBL; BC031381; AAH31381.1; -;  
 DR EMBL; AK035428; BAC29063.1; -;  
 DR MGD; MGI:1322082; 2210413P12RIK.  
 DR InterPro; IPR005805; Rieseke.  
 DR PROSITE; PS00200; RIESEK\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 553 AA; 59742 MW; 6A474PEA630B4B13 CRC64;

Query Match 90.9%; Score 2602; DB 11; Length 553;  
 Best Local Similarity 90.9%; Pred. No. 1.3e-189;  
 Matches 502; Conservative 14; Mismatches 36; Indels 0; Gaps 0;

QY 1 MVQRLVWSRLRRKRAQLLLVLLTFGLVCLAAAGTTPVPLLVGVGVEEKFMTWVLGIG 60  
 DB 1 MVQRLVWSRLRRKRAQLLLVLLTFGLVCLAAAGTTPVPLLVGVGVEEKFMTWVLGIG 60  
 QY 61 PVGLVVCVPLLSASDHWRGRRRPFITWALSGLILLSLFLIPRAGWLAGLCPDPRPL 120  
 DB 61 PVGLVVCVPLLSASDHWRGRRRPFITWALSGLILLSLFLIPRAGWLAGLCPDPRPL 120



```

121 ELALLILGVGLDFCGVCFPTLEALLSDFRDPDHCRCQAVSVYAFMISLGGCLGYLLPA 180
121 ELALLILGVGLDFCGVCFPTLEALLSDFRDPDHCRCQAVSVYAFMISLGGCLGYLLPA 180
181 IDWDTMALAPYLGQOECLFGLLTLIFLTCVAATLLVAEALGPTSPAGLSAPSLSPH 240
181 IDWDTSLAPYLGQOECLFGLLTLIFLTCVAATLLVAEALGPTSPAGLSAPSLSPH 240
241 CCPCRAFLAFNLCALLPRHLQCCRMPTLRRLRFLVAELCSWMAALMTFTFYDFVGEGL 300
241 CCPCRVGLAFNLTGLFRLQCCRMPTLRRLRFLVAELCSWMAALMTFTFYDFVGEGL 300
301 YQGVPRAPGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQFGRFTRVYLASVA 360
301 YQGVPRAPGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQFGRFTRVYLASVA 360
361 APVAAAGATCLSHSVAVVTTASAAITGTFPSALQILPYTLASLYHREKQVFLPKYRGDTG 420
361 TFPVAAAATCLSHSVAVVTTASAAITGTFPSALQILPYTLASLYHREKQVFLPKYRGDTG 420
421 ASSRDSLMTSFLPGPKPGAPPNGHVGAGGSLPPPPALCGASACDVSVRVVGEPTA 480
421 SSSEDSOTTSFLPGPKPGAPPNGHVGAGGSGGILLAPPPALCGASACDVSMRVVGEPTA 480
481 RVVPGRGICLDLALDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAIYPATQ 540
481 RVVPGRGICLDLALDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAIYPATQ 540
541 VVFDKSLDIAKYS 552
541 VVFDKNDIAKYS 552

RESULT 4
ID Q95KCS PRELIMINARY; PRT; 501 AA.
AC Q95KCS;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Hypothetical 53.4 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Osada N., Hida M., Kuwada J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062977; BAB60745.1;
KW Hypothetical protein.
SQ SEQUENCE 501 AA; 53447 MW; 8C554BBD04EE0470 CRC64;

Query Match 88.5%; Score 2531; DB 6; Length 501;
Best Local Similarity 97.8%; Pred. No. 2.8e-184;
Matches 489; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 53 MTWVLGIPVLGVPLVPLGASDHWGRGRRRPTWALSGLILSLFLIPRAGWLAGL 112
DB 1 MTWVLGIPVLGVPLVPLGASDHWGRGRRRPTWALSGLILSLFLIPRAGWLAGL 60
QY 113 LCPDPRPLEALLILGVGLDFCGVCFPTLEALLSDFRDPDHCRCQAVSVYAFMISLGG 172
DB 61 LCPDPRPLEALLILGVGLDFCGVCFPTLEALLSDFRDPDHCRCQAVSVYAFMISLGG 120
QY 173 CLGYLLPAIDWDTMALAPYLGQOECLFGLLTLIFLTCVAATLLVAEALGPTSPAGL 232

```

```

121 CLGYLLPAIDWDTMALAPYLGQOECLFGLLTLIFLTCVAATLLVAEALGPTSPAGL 180
233 SAPSLPHCCPCRAFLAFNLCALLPRHLQCCRMPTLRRLRFLVAELCSWMAALMTFTFY 292
181 SAPSLPHCCPCRAFLAFNLCALLPRHLQCCRMPTLRRLRFLVAELCSWMAALMTFTFY 240
293 TDFVGEGLYQGVPRAPGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQFGR 352
241 TDFVGEGLYQGVPRAPGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQFGR 300
353 AVYLASVAAPVAAAGATCLSHSVAVVTTASAAITGTFPSALQILPYTLASLYHREKQVFLP 412
301 AVYLASVAAPVAAAGATCLSHSVAVVTTASAAITGTFPSALQILPYTLASLYHREKQVFLP 360
413 KYRGDTGGASSEDLSMTSFLPGPKPGAPPNGHVGAGGSLPPPPALCGASACDVSVRV 472
361 KYRGDTGGASSEDLSMTSFLPGPKPGAPPNGHVGAGGSLPPPPALCGASACDVSVRV 420
473 VUGEPTEARVVPGRGICLDLALDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGL 532
421 VUGEPTEARVVPGRGICLDLALDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGL 480
533 VAIYPATQVVFDDKSLDIAKYS 552
481 VAIYPATQVVFDDKSLDIAKYS 500

RESULT 5
ID Q8K252 PRELIMINARY; PRT; 450 AA.
AC Q8K252;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein (Fragment).
GN 2210413P12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034084; AAH34084.1;
DR MGD; MGI:1922082; 2210413P12RIK.
DR InterPro; IPR005805; Rieseke.
DR PROSITE; PS00200; RIESEKE_2; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 450 AA; 48261 MW; 72C703AEC131302C CRC64;

Query Match 73.4%; Score 2099; DB 11; Length 450;
Best Local Similarity 90.0%; Pred. No. 1.7e-151;
Matches 404; Conservative 12; Mismatches 33; Indels 0; Gaps 0;

QY 104 PRAGWLAGLCCPDPRPLEALLILGVGLDFCGVCFPTLEALLSDFRDPDHCRCQAVSV 163
DB 1 PRAGWLAGLCCPDPRPLEALLILGVGLDFCGVCFPTLEALLSDFRDPDHCRCQAVSV 60
QY 164 YAFMISLGGCLGYLLPAIDWDTMALAPYLGQOECLFGLLTLIFLTCVAATLLVAEAL 223
DB 61 YAFMISLGGCLGYLLPAIDWDTMALAPYLGQOECLFGLLTLIFLTCVAATLLVAEAL 120
224 GPTSPAGLSAPSLPHCCPCRAFLAFNLCALLPRHLQCCRMPTLRRLRFLVAELCSWMA 283
121 GPTSPAGLSAPSLPHCCPCRAFLAFNLCALLPRHLQCCRMPTLRRLRFLVAELCSWMA 180
284 ALMTFTFYDFVGEGLYQGVPRAPGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMD 343
181 ALMTFTFYDFVGEGLYQGVPRAPGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMD 240
344 RLVRQFGRFTRVYLASVAAPVAAAGATCLSHSVAVVTTASAAITGTFPSALQILPYTLASLY 403

```

```

Db 241 RLKQKTRSVYLASVMTFPVAAATCLSHVVVVTASALTGTFSAQLLYLASLY 300
QY 404 HREKQVFLPKYRGDTGASSEDLSMTSFLPGPKGAPFPNGHVGAGSGLLPPPPALCGA 463
Db 301 HREKQVFLPKYRGDAGSGSGEDSQTTSFLPGPKGALFPNGHVGSGSGILADPPALCGA 360
QY 464 SACDVSRRVVGGPTEARVPGRGICLDLAILDSAFLLSVAFLSGMSVQLSOSVTAY 523
Db 361 SACDVSRRVVGGPTEARVPGRGICLDLAILDSAFLLSVAFLSGMSVQLSHSVTAY 420
QY 524 MVSAAAGLGLVAIYFATQVVPKDKSLAKYS 552
Db 421 MVSAAAGLGLVAIYFATQVVPKDKSLAKYS 449

RESULT 6
QVSV1 PRELIMINARY; PRT; 599 AA.
AC QVSV1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG4484 protein.
GN CG4484.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKLEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brothier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielson A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003552; AAP50310.1;

```

```

DR Flybase; FBgn0035968; CG4484.
SQ SEQUENCE 599 AA; 66057 MW; C5381D334CFBF2EB CRC64;

Query Match 18.2%; Score 519.5; DB 5; Length 599;
Best Local Similarity 23.7%; Pred. No. 2,7e-31;
Matches 150; Conservative 100; Mismatches 201; Indels 183; Gaps 14;

QY 12 RHKQALLAVNLLTFGLVCLAAGITVVPPLLEVGVEEKFMTVLGIGVPLGLVCVPLL 71
Db 46 RTRPFNFRSLAIAWAIEFAYAEYSFVSPILLQIGVDHKKMSWTWGLSFLIGFMSPLL 105
QY 72 GSASDHWGRYGRRRPFIWALSIGILLSLFLIP-----R 105
Db 106 GSISDRCKLRWGRRPFIISLFGIMCGLILVPGKDLGILLGDCAGTYAESALNTSSS 165
QY 106 AGWLAGLLC-----PDRPLELA--LLILGVGLDFCGQVCTPPEALLSDFLRDPDHC 157
Db 166 GGSVAALVSGEATTCPSASDYKFAVILITLGMVLLDFADTCTQTPARTYLLDMC-VPEEQ 224
QY 158 ROAYSVYAFMISIGGCLGYLLPAIDWDTSAALPYLTQECFLGLTLFLTCVAATL-- 215
Db 225 PKAMTMFALFAGFGGTIGYAGVDWEITHISGFMGGNPTVFTLVTIILPAVCYLITVT 284
QY 216 -----LVAEEAALGP-----TEPABG 231
Db 285 FREIPLPLIEQDELRLPSEQAIKKELKKNTIYIQTTOLELQMSDDPKRLALQG 344
QY 232 LSAPLSLP-----HCCPCRLARLAFNLGALLPRLHQLCCRMPTLRLLF 275
Db 345 SYONGYSPAVERKQKQSDLETQSDYDAPVSLKAYLKSI-----NPYSMRMLA 393
QY 276 VAEELCSWALMTFTLYTDFVGEGLYQGVPAEPGTEARRHYDEGVYRMSGLGLFLOCAIS 335
Db 394 LTNLFCWMGHVTCYLYFTDFVGEAVFHGDPTEAPNSEAALNYEAGVRFGCWGMAYAFSC 453
QY 336 LVFSLVMDRLVORFGRVAVYLASVAAPV-----AAGATCLSHSVAVVTASALTGTFFSAL 392
Db 454 SIYLSVTKLMKMGFTKAVYISGMYYGIGLVLGMLPTKMGVLVFTSA---GILYGTI 510
QY 393 QILPYTLASLYHREKQVFLPKYRGDTGASSEDLSMTSFLPGPKGAPFPNGHVGAGSG 452
Db 511 FTVPPFLVARYH-AKNCFSIK-----NG-----532
QY 453 LLPPPPALCGASACDVSRRVVGGTEARVVP---CRGICLDLAILDSAFLLSVAFLS 509
Db 533 -----BIVLPKQARGIGTDAIISVMVFIAQLIVSLUS 564
QY 510 MGSIVQLSOSVTAYMVVSAAGLGLVAIYFATQVVF 543
Db 565 VGPLVSWMDTTCAVLVASTFLSLAIAAMFVLY 598

RESULT 7
QVR10 PRELIMINARY; PRT; 112 AA.
AC QVR10
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 11.4 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon.
RA Strausberg R.;
DR EMBL; BC024519; AAH24519.1;
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 112 AA; 11393 MW; B9D9DE79FEF09FB7 CRC64;

```

```
Query Match 17.9%; Score 512; DB 11; Length 112;
Best Local Similarity 91.9%; Pred. No. 1.6e-31;
Matches 102; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 442 PNGHVGAGSGLLPPPPALCGACACDVSVRVVGEPTPEARVVGPGICLDLAIDLDSAFLL 501
Db 1 PNGHVGSGSSGILAPPALCGACACDVSMRVVGEPTPEARVVGPGICLDLAIDLDSAFLL 60

QY 502 SOVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVALYFATQVVFQKSLAKYS 552
Db 61 SOVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVALYFATQVVFQKNDLAKYS 111

RESULT 8
Q8C204 PRELIMINARY; PRT; 530 AA.
AC Q8C204;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Membrane associated transporter protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RT "Analysis of the mouse transcriptome based on functional annotation of
RL 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR ENBL; AK089938; BAC1003.1; -.
SQ SEQUENCE 530 AA; 57935 MW; 426998973B7D6420 CRC64;

Query Match 16.9%; Score 484.5; DB 11; Length 530;
Best Local Similarity 26.1%; Pred. No. 1.1e-28;
Matches 146; Conservative 84; Mismatches 229; Indels 101; Gaps 12;

QY 17 QLLVNLFTGLVLCVLAAGITYVPPILLVEGVEKFTMTVLGPGVGLVCPVLLGSASD 76
Db 34 RLVMHSMAMFGREFCYAVEAYVTPGLSVGLPKSLYMWLLSPILGLFQPVVGSASD 93

QY 77 HWRGRRRRPFTWALSGLTLLSLFLIPRAGWLAGLCCDPRP---LEALLILGVLLD 133
Db 94 HCRARGRRRPYLTLLAIMLGMALYLNGLDGVSVLVANPROKLIWASITMVGVLFD 153

QY 134 FCQVQVFTPLEALLSDFRDPDRCQAYSVYAPMISLGGCLGYLLPAIDWDTLSALAPYL 193
Db 154 FSNDFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGALGYLIGRIDWVHLDIGLLG 212

QY 194 TQEECLFGLLTLLIFLCVAATLLVAEEAL--GPTPEP-----ABGLSAPSLSPHCCPC 244
Db 213 TBFQVFFFSSALVLLICFTHLCSIEAPLRDAATDPPSQODPQGSLSASGMHEY---- 268

QY 245 RARLAPNLGA-----LLPLHLQCCRMPTLRLLFVABLCSW 282
Db 269 GSIEKVKNGGADTEQPVQEWKQKPSQSQRTWSMKSLLRALVNMPSHYRCLCVSHLIGW 328

QY 283 MALMTFTLFTVDPVGLGYQVPRAPGTEARRHYDEGRVMSGLGLFQCAISLVESLVW 342
Db 329 TAFLSNMLFTDPMQGIQVHDGPDYGAHNSTEFILYERGVEGVCWGLCINSVSVSYFQ 388

QY 343 DRLVQFGRTRAVLASVAAPFAAGATCLSHSVAVVTAASALTGTFFSALQILPYTLASL 402
Db 389 KAMVSYIGLGLYFMGYLLFGLGTGFIPLFPNVYSVLVCSFMGWSSTLYTVPPNLIAE 448

QY 403 YHREKQVFLPKYRGDTGGASSDSMTSFLPGPKGAPPNGHVAGGSGLLPPPPALCG 462
Db 449 YHREES---KEKGQEA-----PGGPDNQR---GQGV-----DCA 477
```

```
QY 463 ASACDVSVRVVGEPTPEARVVGPGICLDLAIDLDSAFLLSQVAPSLFMGSIQVLSQSVTA 522
Db 478 ALTQVQL-----AQILVGG-----LGLVNMAGSVVV 506

QY 523 YMVSAAGLGLVALYFATQV 542
Db 507 VVITASVSLIGCCFVALFV 526

RESULT 9
Q90Z74 PRELIMINARY; PRT; 576 AA.
AC Q90Z74;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Membrane-associated transporter protein B.
OS Oryzias latipes (Medaka fish) (Japanese ricefish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
OX NCBI_TaxID=8090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HNI;
RX MEDLINE=21372467; PubMed=11479596;
RA Fukunachi S., Shimada A., Shima A.;
RT "Mutations in the gene encoding B, a novel transporter protein, reduce
RL melanin content in medaka.";
RL Nat. Genet. 28:381-385(2001).
DR ENBL; AF332510; AAK7024.1; -.
SQ SEQUENCE 576 AA; 63411 MW; 13A5C9828ABEB894 CRC64;

Query Match 15.9%; Score 456; DB 13; Length 576;
Best Local Similarity 24.0%; Pred. No. 1.7e-26;
Matches 144; Conservative 95; Mismatches 198; Indels 164; Gaps 17;

QY 12 RHRKQALLVNLFTGLVLCVLAAGITYVPPILLVEGVEKFTMTVLGPGVGLVCPVLL 71
Db 56 RSRGRLLHLSHVMFGRFCYAVEAFTVPLLSVGLPRSLYSLVWLISPIGLFLOPII 115

QY 72 GSADHWGRYGRRRPFTWALSGLTLLSLFL-----TPRAGWLAGL 114
Db 116 GSADYCRSSGWRRPYI--LVILGILMVGLSNFLNGDAVSELVSDRSRSTW----- 167

QY 115 PDRPLEALLILGVLLDFCQVCFTPLEALLSDF--FRDPDHCQAYSVYAFMISLGG 172
Db 168 -----AIVVMFGVVLFDFAADFIDGPIKAYLFDVCSYQDKB---RGLHYHALFTGLG 218

QY 173 CLGYLLPAIDWDTLSALAPYLQTEECLEF-----GLLTLLIFLTCV----- 211
Db 219 AGYILVGMWDHGWGLVGLLGSSEYQVIYFFSALTWGVFLVHLFSIPEKPLAKVPSESA 278

QY 212 -AATLLVAEEA---ALG--PTEPAEGLSAPSLSPHCCPC----- 244
Db 279 SSALLRLLGPHSNGYALGKEPVSPVPTSSPIRPSYSALGERPRSFALCEANSVTSS 338

QY 245 -----RARLAPNLGALLPRHLQCCRMPTLRLLFVABLCSMMLMTFTLFTYD 296
Db 339 AKQPKEDQKMTFRS-----LMKAIFNPNHYRFLCISHLGLGAAFLCNMLFTDPM 391

QY 297 GEGLYQVPRAPGTEARRHYDEGRVMSGLGLFQCAISLVPSLVMDRLVQFGRTRAVYL 356
Db 392 GQIVVRGPNPYAHNSTAVITYERGVEGVCWGLCINAVSSALYSYVQRFLLPYIGLGLYF 451

QY 357 ASVAAPFAAGATCLSHSVAVVTAASALTGTFFSALQILPYTLASLYHREKQVFLPKYRG 416
Db 452 MGYFYFGMGTSGLIGLFPPEVIATILCSVFGVMSSTLYTIPNLIAEYQREEEQV-KLEG 510

QY 417 DTGGASSDSMTSFLPGPKGAPPNGHVAGGSGLLPPPPALCGASACDVSVRVVGE 476
Db 511 -----GNESPRGTGM-----DCAALTCMVQL----- 531
```

```
QY 477 PTEARVPGRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAVMVSAAGLGVAVIY 536
Db 532 ---AQIIVGAG-----LGLVNAAGSVIIVVLSASSISLIGCI 566
QY 537 F 537
Db 567 F 567

RESULT 10
Q8K4S3 PRELIMINARY; PRT; 751 AA.
AC Q8K4S3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Proton-associated sugar transporter A.
GN PAST-A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Shimokawa N., Okada J., Haglund K., Dikic I., Koibuchi N., Miura M.;
RT "Past-A, a new class proton-associated sugar transport facilitator in
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB075229; BAB97313.1; --
KW Sugar transport.
SQ SEQUENCE 751 AA; 81751 MW; 5C33E1C2E4A1B2EC CRC64;

Query Match 15.6%; Score 446; DB 11; Length 751;
Best Local Similarity 22.3%; Pred. No. 1.3e-25;
Matches 160; Conservative 79; Mismatches 206; Indels 272; Gaps 15;

QY 17 QLLVNLTLFGLVCLVLAAGTYVPPLLLEVGVEKFTWVLGIGPVILGVLCVPLIGASD 76
Db 86 ELLFNGCILGIEFSYAMETAYVTPVLLQMLPDQSLVFWISPIGLFLQLLQAGWSD 145
QY 77 HWGRGRRRPPFTWALSGLTLLSLELIPRAGLWAGLLCPDP---RPLEALLTLGVGLLD 133
Db 146 RCTSRGRRRPPFLVLAIGALLGLSLIN-GRDIGHALADATATNHWGILLITVCGVLM 204
QY 134 FCQGVCTPLEALLSDFRDPDHCROQVSVYAFMISLGGCLGYLLPAIDWDTSAAPYLG 193
Db 205 FSADSADNPESHAYMMDVCGPVDQR-GLNIHALMAGLGGGFGVVGIIHWKTSFGRALG 263
QY 194 TQECFLGLLTLFLTCVAATLVA-----EEAA 222
Db 264 GOLRVY-IFTATLSTVTVFTLVSPERPLRPLGKRTAMKSPSLPPLPPVLLLEGA 322
QY 223 LGTEP---AEGLSAPSLPHCCP----- 243
Db 323 -GDTLSTATSLYASFSSPSPPLPKYGSFISRDSSLTGINFASFSTNIDSVL 381
QY 244 -----CRALAFNLGALL----- 257
Db 382 IDCFTAGHDNYLALPSSVPRQATSVSPRAPDGYCOERLERREGPLTLGLDGVLRG 441
QY 258 -----PR----- 259
Db 442 SLDTSKPRASGILKRPQTALPDVAGNGPETSRRNRVTFSSQVANILLGVKYESELTG 501
QY 260 -----LHOLCC---RMPTLRRLFAELCSNWMALMTFLFTYDFVGEGLYQGVPR 306
Db 502 SSSQSQPLSLRLCSTIYNMPPVRLNWNHFLGWSLFEGLMLFTYDFMGVGVFQDQPK 561
QY 307 AEPGTEARRHYDEGVKMGSLGLFLOCAISLVFLVMDRLVQRGTTRAVYLASVAAPFVA 366
Db 562 APHASEAYQKNGSVTMCWGMCIYAFSAFYSAILEKLECLSVRTLFIYVLLFGLGT 621
```

```
QY 367 GATCLSHSVAVVTAGAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGCASSED 426
Db 622 GLATLSRNLVVLVSLCTHYGLFSLTCLTLPYSLLCDYYQSK-----KFAQ-----SSADG 671
QY 427 LMTSFLPGPKPGAPFPNGHVHVGAGGGLPPPPALCGASACDVSVRVVVGTEARVVPGR 486
Db 672 TR-----R 674
QY 487 GICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAVMVSAAGLGVAVIYATQVVF 543
Db 675 GMGVDISLLSCQYFLAQILVLSILGP-----LTSVAGSANGVMYFASLVSP 720

RESULT 11
Q8MMT9 PRELIMINARY; PRT; 754 AA.
AC Q8MMT9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE SUCI-sucrose proton symporter.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and Analysis of Chromosome 2 of Dictyostelium.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AC117076; AAM33692.1; --
DR InterPro; IPR005828; Sub transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 754 AA; 82774 MW; 614C473C832ACCE4 CRC64;

Query Match 13.8%; Score 394.5; DB 5; Length 754;
Best Local Similarity 23.5%; Pred. No. 1.1e-21;
Matches 128; Conservative 90; Mismatches 200; Indels 127; Gaps 13;

QY 13 HRKA---QLLVNLTLFGLVCLVLAAGTYVPPLLLEVGVEKFTWVLGIGPVILGVLCV 68
Db 197 HKRTLPFLNLCITCFLGVQFGWALQIAFSTPLFLELGVQEKWVSIWLAGISGLIVQ 256
QY 69 PLLGSASDHWGRGRRRPFIFWALSGLISLISLFLIPRAGLWAGLL--CPDPRPLEALLI 126
Db 257 PLVGVTIDRSECRFGRKPFILIGSVFISGLVLISNAETFGSYFQDSQKKSIAISIAI 316
QY 127 LGVGLLDPCGVCTPLEALLSDFRDPDHCROQVSVYAFMISLGGCLGYLLPAIDWDT 186
Db 317 VGFWILDLSNNAVQAPCRALLVDI-AAPSQSLGSLFSLMLGTGNLLGYMMSI--DLV 373
QY 187 ALAPYLGTQEECLFGLTLTFLTCVAATL-LVAEEAALGPTPAEGLSAPSLPHCCP 245
Db 374 RMVPMKTDTRALFTLSIMVLLFCVVMTLGFTVEQYIRNED-QSVENP----- 422
QY 246 ARLAFNLGALLPRLHQLCCRMPTLRRLFAELCSNWMALMTFLFTYDFVGEGLYQGV 305
Db 423 -----LKTMPKGIVKMPYIQLRCAVQFFSWIGWFSFVLFTTWGVNVFGDP 471
QY 306 RAEPGTEARRHYDEGVKMGSLGLFLOCAISLVFLVMDRLVQRGTTRAVYLASVAAPFVA 365
Db 472 NAEYSDSRILFQDGVKMGSLSLITISSTGIIAVSLIPLFLVKFIDMKYIYIGNL----- 526
QY 366 AGATCLSHSVAVVTAS-----ALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
Db 527 --LQCIFFALFYFVESKISGLLIIASTGIPWALVMTLPFSIVGM----- 568
QY 419 CGASSEDLSMTSFLPGPKPGAPFPNGHVHVGAGGGLPPPPALCGASACDVSVRVVVGTE 478
```

```
Db 569 -GVEDNES-----SG-----577
QY 479 EARVWPGRGICLDLAILDSAFLLSOVAPSLFMSGIYQLSQSVTAYMVSAAGLGLVAIYFA 538
Db 578 -----LNIGTLNIFVVPQWVSLGILGILDLKSGNVVYSLLT---GSVASFFA 623
QY 539 TQVVF 543
Db 624 TLFCF 628

RESULT 12
O65803
ID O65803 PRELIMINARY; PRT; 515 AA.
AC O65803;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Sucrose/H+ symporter.
GN SUT2.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nantaise;
RX MEDLINE=99063785; PubMed=9847123;
RA Shaky R., Sturm A.;
RT "Characterization of source- and sink-specific sucrose/H+ symporters
    from carrot.";
RL Plant Physiol. 118:1473-1480(1998).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; Y16768; CAA76369.1; -.
DR InterPro; IPR005828; Sub transporter.
DR Pfam; PF00083; Suc/H symport.
DR TIGRfam; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 515 AA; 54424 MW; 897B90657C9E243C CRC64;

Query Match 12.1%; Score 347.5; DB 10; Length 515;
Best Local Similarity 25.9%; Pred. No. 2.7e-18;
Matches 131; Conservative 80; Mismatches 215; Indels 79; Gaps 17;

QY 17 QLLVNLTLTFGLEVCCLAAGITVPPLLLEVGVEKEFTMTVLGIGPVGLVGLVCPPLGASD 76
Db 34 KLVLVAIAAGVQFGWALQSLTTPYVQLLGIPIKWAAYIWLCPISGMLVQPIVGYSD 93
QY 77 HWRGRYRRRPFTWALSGLILLSLFLIPRAGWLGL-----LCPDPRLELALLIGV 129
Db 94 HCQSFGRRRPFTASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIG 150
QY 130 GLLDPCQGVCTFLEALLSDLFR-DPDHCRQAYSVAFMISLGGCLGY-----LL 178
Db 151 WILDVANNMLOGPCRALLADLCSGDTREMRSAFYSPFMAVGNILGYAAGSYNNLYKLF 210
QY 179 PAIDWDTLSALAPYLGTOECLFGLLTLFLTCVAATLLVAEEALGPTPEAGLSAPSLS 238
Db 211 PFS--KTHACDLYCANLKSCFIISIALIITVVALSVVRENS--GPPDDADAABEP--- 263
QY 239 PHCCPCRLAFRNIGALLPRHLQCCRMPTRLRLFVAELCSWMALMTFTLYTDFVGE 298
Db 264 ----PSSGKIPV--FGEILGALKOL----PRMILLIITVCLNIAWFFILFDTDMGR 313
QY 299 GLYQGVPAEPGTEARRHYDEGVRMGSLGLFQCAISLVFSLVMDRLVQRFTRAVYLAS 358
Db 314 EYGGT--AGQG-----KLVDQGVRAAGALLNSVVLGTSIAVEYLVRGVGVVKI-LWG 366
QY 359 VAAPFVAAGATCL-----SHSV-----AVVTASA-----ALTGTFSAQLIL 395
```

```
Db 367 FVNFILAIGLVMTVVVSKVAQHREHSANGQLPPSAGVKAGALSILGILPLSITYSI 426
QY 396 PYTLASLHREKQVFLPKYRGDTGGASSEDLSMTSFLPGKPG-----APFPNGHVCA 448
Db 427 PFALASTYSSSGAGGGLSLGLVNLNIAIVVPQMVIVSLAGPFDLSLFGGGLNLPAPVVCALSA 486
QY 449 GSGGLLP-----PPPALCCGASACDVS 469
Db 487 AISGVLAIVLLPKPSKDAASKLSLS 511

RESULT 13
Q9FNR6
ID Q9FNR6 PRELIMINARY; PRT; 515 AA.
AC Q9FNR6;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Sucrose/proton symporter.
GN SUT2.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nantaise;
RA Sturm A.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DDJB databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nantaise;
RA Shaky R.;
RL Thesis (2000), Department of Department of Botany,
    University of Basel, Basel, Switzerland.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AJ303199; CAC19689.1; -.
DR InterPro; IPR005828; Sub transporter.
DR Pfam; PF00083; sugar tr; 1.
DR TIGRfam; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 515 AA; 54425 MW; 9580978C726523DC CRC64;

Query Match 12.1%; Score 347.5; DB 10; Length 515;
Best Local Similarity 25.9%; Pred. No. 2.7e-18;
Matches 131; Conservative 80; Mismatches 215; Indels 79; Gaps 17;

QY 17 QLLVNLTLTFGLEVCCLAAGITVPPLLLEVGVEKEFTMTVLGIGPVGLVGLVCPPLGASD 76
Db 34 KLVLVAIAAGVQFGWALQSLTTPYVQLLGIPIKWAAYIWLCPISGMLVQPIVGYSD 93
QY 77 HWRGRYRRRPFTWALSGLILLSLFLIPRAGWLGL-----LCPDPRLELALLIGV 129
Db 94 HCQSFGRRRPFTASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIG 150
QY 130 GLLDPCQGVCTFLEALLSDLFR-DPDHCRQAYSVAFMISLGGCLGY-----LL 178
Db 151 WILDVANNMLOGPCRALLADLCSGDTREMRSAFYSPFMAVGNILGYAAGSYNNLYKLF 210
QY 179 PAIDWDTLSALAPYLGTOECLFGLLTLFLTCVAATLLVAEEALGPTPEAGLSAPSLS 238
Db 211 PFS--KTHACDLYCANLKSCFIISIALIITVVALSVVRENS--GPPDDADAABEP--- 263
QY 239 PHCCPCRLAFRNIGALLPRHLQCCRMPTRLRLFVAELCSWMALMTFTLYTDFVGE 298
Db 264 ----PSSGKIPV--FGEILGALKOL----PRMILLIITVCLNIAWFFILFDTDMGR 313
QY 299 GLYQGVPAEPGTEARRHYDEGVRMGSLGLFQCAISLVFSLVMDRLVQRFTRAVYLAS 358
Db 314 EYGGT--AGKG-----KLVDQGVRAAGALLNSVVLGTSIAVEYLVRGVGVVKI-LWG 366
```

```
QY 359 VAAPFVAAGATCL-----SHSV-----AVVTASA-----ALTGTFPSALQIL 395
Db 367 FNFILAIAGLMTVVVVKVAQHOREHSANGQLPPSAGVAGALSLSILGIPLSITYSI 426
QY 396 PYTLASLYHREKQVFLPKYRGDTGGASSEDLSMTSLPQPKPG-----APFPNGHVGA 448
Db 427 PFALASIYSGSGAGGGLSLGVNLAIIVPQMIVSVLAGPFDLSFGGGLNLPARVVGAI 486
QY 449 GSGGLLP-----PPPALCGASACDVS 469
Db 487 AISGVLAIVLLPKPSKDAASKLSLS 511

RESULT 14
Q9SQK6 PRELIMINARY; PRT; 501 AA.
AC Q9SQK6;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Putative sucrose transporter.
GN VVSUC11.
OS Vitis vinifera (Grape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC Vitis.
OX NCBI_TaxID=29760;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Shiraz; TISSUE=Fruit;
RA Davies C., Wolf T., Robinson S.P.;
RT "Three putative sucrose transporters are differentially expressed in
RT grapevine tissues.";
RL Plant Sci. 147:93-100(1999).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AF021808; AAF08329.1; -.
DR InterPro: IPR005828; Sub transporter.
DR InterPro: IPR005989; Suc/H_symport.
DR Pfam: PF00083; sugar tr; 1.
DR TIGRFAMs: TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 501 AA; 53938 MW; 4D0D4DE2EF2F4BA8 CRC64;
```

```
Query Match 12.1%; Score 346.5; DB 10; Length 501;
Best Local Similarity 27.0%; Pred. No. 3.1e-18;
Matches 153; Conservative 83; Mismatches 197; Indels 133; Gaps 22;

QY 17 QLLLVNLLTFGLVCLAAAGITVPPLLLEVGVEEKFTMTVLGIGPVGLVCVPLIGSASD 76
Db 29 RLLRASVACGIQFGWALQSLTTPVQELGIPHAMSSIIWLCGLSGLLVQPLVGHLS 88
QY 77 HWRGRRRRPFIWALSGLLSLFLIPRAGWLAGLL--CPDPRPLEALLILGVGLDF 134
Db 89 RCNSRFGRRRPFIVAGATSIWVAVLIIGPSADIGLLGDGADRPRPRAVATFVVGFWLLDV 148
QY 135 CGQVCTPPEALLSDLPDPH--CRQAYSVYAFMISLGCGLYLPA-----IDWDT 186
Db 149 ANNVTOGPCRALLADL--TEKHRRTRVANYAFSLFIAVGNVLGATGSGVGFRIWFMTS 207
QY 187 ALAPYLGTQECI---FGLTLTILFTCVATLLVAEEAALGPTPEAGLSAPLSPHCC 242
Db 208 TSS-----CNADCANLKSFAFLDIIF--IAITYISITAA-----QELPLSSSRSTHIS 255
QY 243 -----PCRALPARNLIGALLPRHLQCCRMPTLRR-----LFVAELCSWMLMTFT 289
Db 256 EEMAESTHAQEA-----LWELF-----GTLRYLSGSIWILFVTAL--TWIGLLPFL 301
QY 290 LFTYDFVGEGLYQGVPAEPTGARHYDEGVNMGSLGLFLQCAISLVFSLVMDRLVQRF 349
Db 302 LFTDWMGRIYGGKP-----NEGQNTYGVRMGALGMLNSVILGITSVLMEKLCRW 355
QY 350 GTRAVVLAASVAAPFAAGATCLSHSVAVVTASAALTGTFPSALQILPYTLASLYHREKQV 409
```

```
Db 356 GAGFVW-----GLSNILMSLCFLMLLILS-----AVVGHMD----- 386
QY 410 FLPKYRGDTGGASSEDLSMTSLPQPKPGAPPNHVGAGSGGLLPPPPALCGASACDVS 469
Db 387 -----FLGHDLPP-----PSGVWIA-----ALIVFSILGIP 411
QY 470 VVVVVVGEPP--TEARVVP--GRGICICLDLALDSAFLLSOPVAPSLFMGSIVQL-----SQSV 520
Db 412 LAITYSVFVALISTRIEISGLGQGLSMGVNLAIIVPQIVISLGSGLPQDLFGGNSPSL 471
QY 521 TAYMVSAAAGLGLVAIYFATQVVFDSK 546
Db 472 AVAAVAAPASGLVAIILAIPTSSADKS 497

RESULT 15
Q9SLN7 PRELIMINARY; PRT; 515 AA.
AC Q9SLN7;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Sucrose transporter protein.
GN CSUT.
OS Daucus carota (Carrot).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridaceae; campanulids; Apiales; Apiaceae; Daucus.
OX NCBI_TaxID=4039;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Root;
RA Wu N., Diao F., Zhang L., Huang M.;
RT "Sucrose transporter protein.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL: AB036758; BAA89458.1; -.
DR InterPro: IPR005828; Sub transporter.
DR InterPro: IPR005989; Suc/H_symport.
DR Pfam: PF00083; sugar tr; 1.
DR TIGRFAMs: TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 515 AA; 54523 MW; 2B548D1D9DFF51AC CRC64;
```

```
Query Match 12.1%; Score 346.5; DB 10; Length 515;
Best Local Similarity 27.2%; Pred. No. 3.2e-18;
Matches 137; Conservative 76; Mismatches 215; Indels 75; Gaps 19;

QY 17 QLLLVNLLTFGLVCLAAAGITVPPLLLEVGVEEKFTMTVLGIGPVGLVCVPLIGSASD 76
Db 34 KULVVAATAAGVQFGWALQSLTTPVQELGIPHAMSSIIWLCGLSGLLVQPLVGVYED 93
QY 77 HWRGRRRRPFIWALSGLLSLFLIPRA---GWLAG--LLCPDPRPLEALLILGVGLL 132
Db 94 HCQSSFGRRRPFIVAGATSIWVAVLIIGPSADIGYKAGDDMSKTLKPRAVTFVIGFWIL 153
QY 133 DFCQGVCTPPEALLSDLPFR--DPDCHCRQAYSVYAFMISLGCGLY-----LLPAI 181
Db 154 DVANNMLOGPCRALLADLCSGDTRRMRSSANAFYSFPMVGNILGYAAGSYNNLYKLFPFS 213
QY 182 DWDTSALAPYLGTQECI--LFGLLTILFTCVATLLVAEEAALGPTPEAGLSAPLSPH 240
Db 214 --KTHACDLYCANLKSFCFIISIALIITVVA--LSVREKQWSP--DDADADEP----- 263
QY 241 CCPCRALPARNLIGALLPRHLQCCRMPTLRRLLFVAELCSWMLMTFTLFTYDFVGEGL 300
Db 264 --PSSGKIPV--FGELLGALKDL---PRPMLLLIIVTCLNNIAWPPFLIFDTDMWGRI 315
QY 301 YQGVPAEPTGARHYDEGVNMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVILASVA 360
Db 316 YGGT--AGKG-----KLYDQGVRAAGSLGSLNSVILGITSIAVEYLVRGVGGVKI--LWGV 368
QY 361 AFPVAAGATCL-----SHSV-----AVVTASA-----ALTGTFPSALQILPY 397
```

369	NEILAIGLMTVTWVSKVAHQREHSANGQLLPPSAGVKGALSLFSLGIPUSITYSIPF	428
Db	:	:
398	TLASLYHREKQVFLPKYRGDTGGASSEDLSMTSFLPGPKPG-----APFPNGHVHGAGG	450
Qy	:	:
429	ALASIIYSGSGAGGCLSLGLVNLAIVVPQIVSVLAGPFDLSFGGNGLPAFVVGAISRAI	488
Db	:	:
451	SGLLP-----PPPALCGSACDVS	469
Qy	:	:
489	SGVLAIVLLPKPSKDAASKLSLS	511
Db	:	:

Search completed: December 3, 2003, 17:20:41  
Job time : 42 secs

---

**THIS PAGE BLANK (USPTO)**

---

**THIS PAGE BLANK (USPTO)**





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:19:21 ; Search time 47 seconds  
(without alignments)  
1867.570 Million cell updates/sec

Title: US-09-593-793a-113

Perfect score: 553

Sequence: 1 MVQRLVMSRLRRKKAQLLL.....AIYFATQVDFKSLAKYSA 553

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1107863 seqs, 158726573 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

A Geneseq 19Jun03:\*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
- 24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	553	100.0	553	19	Amino acid encoded
2	553	100.0	553	19	Prostate tumour sp
3	553	100.0	553	21	Protein encoded by
4	553	100.0	553	21	Human immunogenic
5	553	100.0	553	22	Human prostate cDN
6	553	100.0	553	22	Human prostate tum
7	553	100.0	553	22	Human prostate-spe
8	553	100.0	553	22	Human prostate-spe
9	553	100.0	553	22	Human P501S invent

10	553	100.0	553	22	AA74800	Prostate tumour an
11	553	100.0	553	23	ABG94411	Human prostate tum
12	553	100.0	553	23	ABG76665	Prostate tumour pr
13	553	100.0	553	23	ABG77575	Human mast cell re
14	553	100.0	553	23	ABG61900	Prostate cancer-as
15	553	100.0	553	23	ABG5222	Human L1-12 protei
16	553	100.0	553	23	AAU10324	Human PROST 03. H
17	553	100.0	553	23	AAU82643	Human breast tumou
18	553	100.0	553	24	ABU71653	Prostate cancer sp
19	553	100.0	1079	22	ABU74830	Prostate tumour an
20	553	100.0	1079	24	ABU71860	Prostate specific
21	501	90.6	710	23	AAW50661	Thioredoxin-ubiqui
22	499	90.2	595	22	AAW01318	Alpha prepro-P501S
23	320	57.9	359	24	ABU71887	Human prostate spe
24	314	56.8	530	23	AAW50662	Human prostate spe
25	263	47.6	371	22	AAU69875	Thioredoxin-ubiqui
26	263	47.6	371	22	AAW01230	Human prostate cDN
27	263	47.6	371	22	AAW01362	P53S splice varia
28	263	47.6	371	23	ABG64105	Human albumin fusi
29	263	47.6	371	23	ABG95335	Human P553S splice
30	263	47.6	371	24	ABU71766	Prostate cancer as
31	263	47.6	400	22	AAU69907	Human prostate pro
32	263	47.6	400	22	AAW01262	Human prostate-E2 cons
33	263	47.6	400	23	ABG95367	Rai2-P501S-E2 cons
34	263	47.6	400	24	ABU71798	Prostate cancer as
35	255	46.1	255	20	AAW85068	Protein encoded by
36	255	46.1	255	21	AAW29268	Human prostate-rel
37	255	46.1	255	22	AAU04205	Prostate-specific
38	255	46.1	255	23	AAO19084	Human prostate-spe
39	206	37.3	231	21	AAW54369	Amino acid sequenc
40	187	33.8	326	23	ABW77571	Human mast cell re
41	161	29.1	252	22	AAE01423	Human secreted pro
42	87	15.7	123	22	AAU69873	Human prostate cDN
43	87	15.7	123	22	AAW01228	P53S splice varia
44	87	15.7	123	23	ABG95333	Human P553S splice
45	87	15.7	123	24	ABU71764	Prostate cancer as

#### ALIGNMENTS

#### RESULT 1

AAW71869

ID AAW71869 standard; Protein; 553 AA.

AC AAW71869;

06-JAN-1999 (first entry)

Amino acid encoded by prostate tumour clone L1-12.

Prostate; cancer; tumour; vaccine; immunogen; clone.

Homo sapiens.

WO9837093-A2.

27-AUG-1998.

25-FEB-1998; 98WO-US03492.

09-FEB-1998; 98US-0020956.

25-FEB-1997; 97US-0806099.

01-AUG-1997; 97US-0904804.

(CORI-) CORIXA CORP.

Dillon DC, Xu J;

WPI, 1998-609886/51.

N-PSDB; AAV61201.

Polypeptides comprising immunogenic portions of prostate proteins -

PT used in a vaccine for the treatment of prostate cancer  
 XX  
 PS Example 1; Page 82-84; 130pp; English.

XX The present sequence is an immunogenic portion of a prostate tumour  
 CC protein. The immunogen, or the DNA encoding it, can be used as a  
 CC vaccine for the treatment of prostate cancer. The immunogen was  
 CC isolated from a prostate tumour cDNA library obtained by subtracting  
 CC a prostate tumour cDNA expression library with a normal tissue cDNA  
 XX library.

XX SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 19; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVORLWVSRLRRHKAQQLLVNLLTFGLEVCCLAAGITYVPPLLELVGVBEKFTMWLGIG 60  
 DB 1 MVORLWVSRLRRHKAQQLLVNLLTFGLEVCCLAAGITYVPPLLELVGVBEKFTMWLGIG 60  
 QY 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWDTLSALAPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWDTLSALAPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLARLAFRNIGALLPRHQLCCRMPTLRRLFVAELCSNMALMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLARLAFRNIGALLPRHQLCCRMPTLRRLFVAELCSNMALMTFTLYTDFVGEGL 300  
 QY 301 YQGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 AFPVAAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTGA 480  
 DB 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTGA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVMSAAGLGLVAYIFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVMSAAGLGLVAYIFATQ 540

RESULT 2

AAW69385

ID AAW69385 standard; Protein; 553 AA.

XX

AC AAW69385;

XX

DT 25-MAR-2003 (updated)

XX 08-DEC-1998 (first entry)

XX

DE Prostate tumour specific gene clone L1-12 protein.

XX Prostate tumour specific gene; human; prostate cancer; detection;

KW therapy.

XX

OS Homo sapiens.

XX WO9837418-A2.  
 XX  
 PD 27-AUG-1998.  
 XX  
 PF 25-FEB-1998; 98WO-US03690.  
 XX  
 PR 25-FEB-1997; 97US-0806596.  
 PR 01-AUG-1997; 97US-0904809.  
 PR 09-FEB-1998; 98US-0020747.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Dillon DC, Xu J;  
 XX  
 DR WPI; 1998-480805/41.  
 DR N-PSDB; AAV58586.  
 XX  
 PT Novel human prostate specific tumour protein and fragments - useful  
 PT for detecting and treating prostate cancer  
 XX  
 PS Example 1; Page 87-89; 141pp; English.  
 XX  
 CC This sequence is encoded by a human prostate tumour specific gene, and  
 CC can be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprises contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC this protein sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect;  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 CC (updated on 25-MAR-2003 to correct PR field.)  
 XX  
 SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 19; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVORLWVSRLRRHKAQQLLVNLLTFGLEVCCLAAGITYVPPLLELVGVBEKFTMWLGIG 60  
 DB 1 MVORLWVSRLRRHKAQQLLVNLLTFGLEVCCLAAGITYVPPLLELVGVBEKFTMWLGIG 60  
 QY 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVGLVCVPLLGASDHWGRYGRRRPFIWALSGLILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWDTLSALAPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWDTLSALAPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLARLAFRNIGALLPRHQLCCRMPTLRRLFVAELCSNMALMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLARLAFRNIGALLPRHQLCCRMPTLRRLFVAELCSNMALMTFTLYTDFVGEGL 300  
 QY 301 YQGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 DB 301 YQGVPAEPTGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 AFPVAAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTGA 480  
 DB 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTGA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVMSAAGLGLVAYIFATQ 540

Db 481 RVVPGRGICLDLAIDSAFLISQVAPSLFMGSIIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
XX AC  
XX AC  
XX AC  
DT 07-FEB-2001 (first entry)  
XX  
XX Protein encoded by human breast tumour cDNA clone P501S.  
XX Human; breast tumour antigen; cytostatic; immunotherapy;  
KW breast cancer; vaccine.  
XX  
XX Homo sapiens.  
XX WO200061756-A2.  
XX  
XX 19-OCT-2000.  
XX  
XX 10-APR-2000; 2000WO-US09688.  
XX  
XX 09-APR-1999; 99US-0288950.  
XX 02-JUL-1999; 99US-0346327.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Reed SG, Xu J, Dillon DC;  
XX WPI; 2000-638568/61.  
XX N-PSDB; AAC79473.  
XX  
XX A novel isolated polypeptide comprising an immunogenic portion of a  
PT breast cancer protein useful in the detection and treatment of breast  
PT cancer -  
XX  
XX Claim 2; Page 92-93; 95pp; English.  
XX  
XX The present sequence is encoded by a cDNA sequence which was isolated  
CC from a breast tumour cDNA library. It is provided in a specification  
CC relating to compounds for immunotherapy and diagnosis of breast cancer.  
CC Breast tumour antigens and the polynucleotides that encode them may be  
CC used in the production of a pharmaceutical composition to be used in the  
CC treatment of breast cancer. Proliferated T cells and incubated antigen  
CC presenting cells are also required. The polypeptides and polynucleotides  
CC may also be used to produce a vaccine.  
XX  
XX Sequence 553 AA;  
SQ  
Query Match 100.0%; Score 553; DB 21; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVQRLWVSRLLRRKKAQLLVNLLTFGLEVCLAAAGITYVPPLLEVGVEKFTMWLGIG 60  
Db 1 MVQRLWVSRLLRRKKAQLLVNLLTFGLEVCLAAAGITYVPPLLEVGVEKFTMWLGIG 60  
QY 61 PVLGLVCVPLIGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
Db 61 PVLGLVCVPLIGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120  
QY 121 ELALLILGVLLDFCCGVCTPTEALLSDLFDPDHCROAYSVYAFWISLGGCLGLLPA 180  
Db 121 ELALLILGVLLDFCCGVCTPTEALLSDLFDPDHCROAYSVYAFWISLGGCLGLLPA 180  
QY 181 IDWDTSAALPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240

Db 181 IDWDTSAALPYLGTQBECLFGLLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
QY 241 CCPCRLARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSNMALMTFTLFYTDFFVGEGL 300  
Db 241 CCPCRLARLAFRNIGALLPRHLQCCRMPTLRRLFVAELCSNMALMTFTLFYTDFFVGEGL 300  
QY 301 YQGVPRAEPTGEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
Db 301 YQGVPRAEPTGEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
QY 361 APFVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 APFVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPTTEA 480  
Db 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPTTEA 480  
QY 481 RVVPGRGICLDLAIDSAFLISQVAPSLFMGSIIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540  
Db 481 RVVPGRGICLDLAIDSAFLISQVAPSLFMGSIIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553  
RESULT 4  
AAB28527  
ID AAB28527 standard; Protein; 553 AA.  
XX AC  
XX AC  
XX AC  
DT 13-JUN-2000 (first entry)  
XX  
XX Human immunogenic prostate tumour protein sequence SEQ ID NO:113.  
DE DE  
XX Human; prostate cancer; diagnosis; tumour; gene therapy; detection;  
KW immunogenic; cytostatic; vaccine.  
XX  
XX Homo sapiens.  
XX  
XX WO200004149-A2.  
XX  
XX 27-JAN-2000.  
XX  
XX 14-JUL-1999; 99WO-US15838.  
XX  
XX 14-JUL-1998; 98US-0115453.  
XX 14-JUL-1998; 98US-0116134.  
XX 23-SEP-1998; 98US-0159812.  
XX 23-SEP-1998; 98US-0159822.  
XX 15-JAN-1999; 99US-0232149.  
XX 15-JAN-1999; 99US-0232880.  
XX 09-APR-1999; 99US-0288946.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Dillon DC, Harlocker SL, Yugu J, Xu J, Mitcham JL;  
XX WPI; 2000-171268/15.  
XX  
XX New polypeptide useful for treating and diagnosing prostate cancer  
PT comprises an immunogenic portion of prostate tumor protein -  
PT  
XX  
XX Claim 3; Page 138-139; 263pp; English.  
XX  
XX The present invention describes isolated polypeptides, comprising an  
CC immunogenic portion of a prostate tumour protein (PTP). The polypeptides  
CC and polynucleotides encoding them have cytostatic activity and can be  
CC used in vaccines and in gene therapy. The polypeptides and  
CC polynucleotides encoding them, antigen presenting cells which express

CC the polypeptides, antibodies against the polypeptides and vaccines  
 CC comprising them can be used for inhibiting the development of prostate  
 CC cancer in a patient. The polypeptides can be used to generate antibodies  
 CC or anti-idiotype antibodies for passive immuno therapy. A portion of  
 CC the polynucleotides encoding the polypeptides can be used as a probe or  
 CC to modulate the expression of the polypeptides. AAA06241 to AAA06691 and  
 CC AAY82000 to AAY82020 represent sequences used in the exemplification of  
 CC the present invention.

XX  
 SQ Sequence 553 AA;  
 Query Match 100.0%; Score 553; DB 21; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLLVNLTLFGLVCLAAAGTTPVPLLELVGVVEKFTWVLGIG 60  
 DB 1 MVQRLWVSRLLRHRKAQLLLVNLTLFGLVCLAAAGTTPVPLLELVGVVEKFTWVLGIG 60  
 QY 61 PVLGLVCPVLLGSASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVLGLVCPVLLGSASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLILVGLLDFCGQVCFTEALLSDFRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILVGLLDFCGQVCFTEALLSDFRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWDTLSALAPYLGTQECFLGILLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWDTLSALAPYLGTQECFLGILLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLARAFNLGALLPRHLQCCRMPTLRLLFVAELCSNMALMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLARAFNLGALLPRHLQCCRMPTLRLLFVAELCSNMALMTFTLYTDFVGEGL 300  
 QY 301 YQGVPAEPGTEARRHYDEGVGMGSLGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
 DB 301 YQGVPAEPGTEARRHYDEGVGMGSLGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAQLPPTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAQLPPTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGSLPPPPALCGASACDVSRVVVGEPTA 480  
 DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGSLPPPPALCGASACDVSRVVVGEPTA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIPATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAYIPATQ 540  
 QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

RESULT 5  
 AAU69763  
 ID AAU69763 standard; Protein; 553 AA.  
 XX  
 AC AAU69763;  
 XX  
 XX  
 DT 30-JAN-2002 (first entry)  
 XX  
 DE Human prostate cDNA encoded protein #3.  
 XX  
 KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.  
 OS Homo sapiens.  
 XX  
 PN W0200173032-A2.  
 XX  
 PD 04-OCT-2001.

XX  
 PF 27-MAR-2001; 2001WO-US09919.  
 XX  
 XX 27-MAR-2000; 2000US-0536857.  
 PR 09-MAY-2000; 2000US-0568100.  
 PR 12-MAY-2000; 2000US-0570737.  
 PR 13-JUN-2000; 2000US-0593793.  
 PR 17-JUN-2000; 2000US-0605783.  
 PR 10-AUG-2000; 2000US-0636215.  
 PR 29-AUG-2000; 2000US-0651236.  
 PR 06-SEP-2000; 2000US-0657279.  
 PR 02-OCT-2000; 2000US-0679426.  
 PR 10-OCT-2000; 2000US-0685166.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedwick TS, Carter D;  
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;  
 XX WPI; 2001-639232/73.  
 DR N-PSDB; AAS63557.  
 XX  
 PT New human prostate-specific polypeptides and polynucleotides useful for  
 the diagnosis and treatment of cancer, especially prostate cancer -  
 XX  
 PS Claim 2; Page 269-270; 579pp; English.  
 XX  
 CC The invention relates to isolated prostate-specific  
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,  
 CC antibodies raised against the polypeptides for antigenic epitopes  
 CC derived from them) and antigen-presenting cells expressing the  
 CC polypeptides. The antibodies are useful for detecting the presence of  
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and  
 CC the antigen-presenting cells are useful for stimulating and/or expanding  
 CC T cells specific for a tumour protein, and for inhibiting the development  
 CC of cancer especially prostate cancer. Compositions comprising an immune  
 CC polynucleotide and/or polypeptide are useful for stimulating an immune  
 CC response, and for treating cancer. The oligonucleotide is useful for  
 CC detecting cancer. The present sequence is a prostate specific  
 CC polypeptide of the invention.

XX  
 SQ Sequence 553 AA;  
 Query Match 100.0%; Score 553; DB 22; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLLVNLTLFGLVCLAAAGTTPVPLLELVGVVEKFTWVLGIG 60  
 DB 1 MVQRLWVSRLLRHRKAQLLLVNLTLFGLVCLAAAGTTPVPLLELVGVVEKFTWVLGIG 60  
 QY 61 PVLGLVCPVLLGSASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVLGLVCPVLLGSASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLILVGLLDFCGQVCFTEALLSDFRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILVGLLDFCGQVCFTEALLSDFRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 QY 181 IDWDTLSALAPYLGTQECFLGILLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWDTLSALAPYLGTQECFLGILLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLARAFNLGALLPRHLQCCRMPTLRLLFVAELCSNMALMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLARAFNLGALLPRHLQCCRMPTLRLLFVAELCSNMALMTFTLYTDFVGEGL 300  
 QY 301 YQGVPAEPGTEARRHYDEGVGMGSLGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
 DB 301 YQGVPAEPGTEARRHYDEGVGMGSLGLFLQCAISLVFSLVMDRLVQRFQTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAALTGFTFSAQLPPTLASLYHREKQVFLPKYRGDTGG 420







```
CC AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQLLNVLLTFGLVCLAAAGITYVPPLLLLEVGVBEKFTMWLGIG 60
Db 1 MVQRLWVSRLLRHRKAQLLNVLLTFGLVCLAAAGITYVPPLLLLEVGVBEKFTMWLGIG 60
QY 61 PVGLVVCVPLLGASDHWRGRRRPFITWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
Db 61 PVGLVVCVPLLGASDHWRGRRRPFITWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
Db 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
Db 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
QY 241 CCPCRLARLAFRNIGALLPRLHQLCCMRPTLRRLFVAELCSWMLMTFTFLFYTDFV 300
Db 241 CCPCRLARLAFRNIGALLPRLHQLCCMRPTLRRLFVAELCSWMLMTFTFLFYTDFV 300
QY 301 YQGVPRAEPTGEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYL 360
Db 301 YQGVPRAEPTGEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYL 360
QY 361 AFPVAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYR 420
Db 361 AFPVAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYR 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNHGVGAGSGLLPPPPALCGASACDVSVRVVV 480
Db 421 ASSEDSLMTSFLPGPKPGAPFNHGVGAGSGLLPPPPALCGASACDVSVRVVV 480
QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLV 540
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLV 540
QY 541 VFDDKSDLAKYSA 553
Db 541 VFDDKSDLAKYSA 553

RESULT 11
ABG94411
ID ABG94411 standard; Protein; 553 AA.
XX
AC ABG94411;
XX
DT 27-NOV-2002 (first entry)
XX.
XX Human prostate tumour protein partial sequence #3.
XX
KW Human; immunogenic; prostate protein; prostate tumour protein;
KW prostate cancer; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
XX US2002090372-A1.
XX
XX 11-JUL-2002.
XX
XX 14-JUL-1998; 98US-0115453.
XX
XX 25-FEB-1997; 97US-0806099.
XX
XX 01-AUG-1997; 97US-0904804.
XX
```

```
PR 09-FEB-1998; 98US-0020956.
PR 25-FEB-1998; 98US-0030607.
XX
XX (XUJ/) XU J.
XX (DILL/) DILLON D C.
XX
XX Xu J, Dillon DC;
XX
XX WPI: 2002-642373/69.
XX N-PSDB; ABS21254.
XX
XX Novel polypeptides useful as vaccines for inhibiting prostate cancer
XX development, comprise an immunogenic portion of prostate protein -
XX
XX Example 1; Page 59-61; 101pp; English.
XX
XX The present invention relates to a new polypeptide comprising an
XX immunogenic portion of a prostate protein. The invention is useful for
XX inhibiting the development of prostate cancer in a patient. The
XX invention is also useful as markers for diagnosing prostate cancer and
XX for monitoring diseases progression in patients. The present amino
XX acid sequence represents a human prostate tumour protein.
XX
XX Sequence 553 AA;
XX
XX Query Match 100.0%; Score 553; DB 23; Length 553;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 MVQRLWVSRLLRHRKAQLLNVLLTFGLVCLAAAGITYVPPLLLLEVGVBEKFTMWLGIG 60
Db 1 MVQRLWVSRLLRHRKAQLLNVLLTFGLVCLAAAGITYVPPLLLLEVGVBEKFTMWLGIG 60
QY 61 PVGLVVCVPLLGASDHWRGRRRPFITWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
Db 61 PVGLVVCVPLLGASDHWRGRRRPFITWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
Db 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
Db 181 IDWDTSAALPYLGTQBECLFGLLTFLITCVAATLLVAEEAALGPTPEAEGLSAPLS 240
QY 241 CCPCRLARLAFRNIGALLPRLHQLCCMRPTLRRLFVAELCSWMLMTFTFLFYTDFV 300
Db 241 CCPCRLARLAFRNIGALLPRLHQLCCMRPTLRRLFVAELCSWMLMTFTFLFYTDFV 300
QY 301 YQGVPRAEPTGEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYL 360
Db 301 YQGVPRAEPTGEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYL 360
QY 361 AFPVAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYR 420
Db 361 AFPVAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYR 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNHGVGAGSGLLPPPPALCGASACDVSVRVVV 480
Db 421 ASSEDSLMTSFLPGPKPGAPFNHGVGAGSGLLPPPPALCGASACDVSVRVVV 480
QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLV 540
Db 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLV 540
QY 541 VFDDKSDLAKYSA 553
Db 541 VFDDKSDLAKYSA 553

RESULT 12
ABG76665
ID ABG76665 standard; Protein; 553 AA.
```



XX ABG76665;  
 AC 05-NOV-2002 (first entry)  
 DT Prostate tumour protein #3.  
 DE Human; prostate tumour; immunotherapy; prostate cancer.  
 XX Homo sapiens.  
 OS US2002081580-A1.  
 PN 27-JUN-2002.  
 PD 25-FEB-1998; 98US-0030606.  
 PF 25-FEB-1997; 97US-0806596.  
 PR 01-AUG-1997; 97US-0904809.  
 PR 09-FEB-1998; 98US-0020747.  
 XX (XUJJ/) XU J.  
 PA (DILL/) DILLON D C.  
 PI Xu J, Dillon DC;  
 FI WPI; 2002-607662/65.  
 DR  
 DT Detecting prostate cancer comprises contacting a sample with an agent  
 PT capable of binding to a polypeptide with an immunogenic portion of a  
 PT prostate protein, oligonucleotide primers or a probe specific for DNA  
 PT encoding the polypeptide -  
 XX Example 1; Page 65-67; 11pp; English.  
 PS The invention relates to a method of detecting prostate cancer by  
 XX contacting a biological sample from a patient with: (a) a binding agent  
 CC that binds to a polypeptide having an immunogenic portion of a prostate  
 CC protein or its variant; (b) 2 oligonucleotide primers, where 1 of the  
 CC oligonucleotides is specific for a DNA encoding the polypeptide of (a);  
 CC or (c) an oligonucleotide probe specific for a DNA molecule encoding the  
 CC polypeptide of (a). The method and polypeptides are useful for  
 CC diagnosing, treating, particularly by immunotherapy, monitoring the  
 CC progression, and inhibiting the development of prostate cancer in a  
 CC patient. The polypeptides may be used to generate antibodies useful for  
 CC the diagnosis and monitoring of prostate cancer. ABG76663-ABG76669  
 CC represent human prostate tumour protein sequences of the invention.  
 XX SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 23; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVORLWVSLRLRHRKAQLLLVLLTFLGLEVLAAGITVPPLLEVGVEKFKMTWVLIG 60  
 DB 1 MVORLWVSLRLRHRKAQLLLVLLTFLGLEVLAAGITVPPLLEVGVEKFKMTWVLIG 60  
 QY 61 PVLGLVCVPLLLGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVLGLVCVPLLLGSASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
 QY 121 ELALLIIGVLLDFCGQVCTFEALISDLFRPDHCRQAYSVYAFWISIGGCGVLLPA 180  
 DB 121 ELALLIIGVLLDFCGQVCTFEALISDLFRPDHCRQAYSVYAFWISIGGCGVLLPA 180  
 QY 181 IDWTSALAPYLGTEBCEFLGLTLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 DB 181 IDWTSALAPYLGTEBCEFLGLTLTFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
 QY 241 CCPCRLAPRNIGALLPRLHQLCCMRPTRLRLFVAELCSWMLMTFTLYTDFVGEGL 300  
 DB 241 CCPCRLAPRNIGALLPRLHQLCCMRPTRLRLFVAELCSWMLMTFTLYTDFVGEGL 300

QY 301 YQGVPRAPCTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
 DB 301 YQGVPRAPCTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
 QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 QY 421 ASSEDSIMTSFLPGPKGAPFPNGHVAGSGGLPPPPALCGASACDVSVRVVVGPTPEA 480  
 DB 421 ASSEDSIMTSFLPGPKGAPFPNGHVAGSGGLPPPPALCGASACDVSVRVVVGPTPEA 480  
 QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSPFMGSIQVLSQSVTAYMWSAAGLGLVAIYPATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSPFMGSIQVLSQSVTAYMWSAAGLGLVAIYPATQ 540  
 QY 541 VFEDKSDLAKYSA 553  
 DB 541 VFEDKSDLAKYSA 553

RESULT 13  
 ABB77575  
 ID ABB77575 standard; Protein; 553 AA.  
 XX ABB77575;  
 AC ABB77575;  
 DT 30-AUG-2002 (first entry)  
 XX Human mast cell related splice variant protein MC14 SEQ ID NO 13.  
 DE Human; mast cell; MC; antiallergic; antiinflammatory; antiasthmatic;  
 XX vasotrophic; dermatological; allergy; hypersensitivity; rhinitis; asthma.  
 OS Homo sapiens.  
 PN WO200246389-A2.  
 PD 13-JUN-2002.  
 XX 07-DEC-2001; 2001WO-US46180.  
 XX 08-DEC-2000; 2000US-251835P.  
 PR 14-MAR-2001; 2001US-275479P.  
 PR 28-MAR-2001; 2001US-279115P.  
 PR 02-APR-2001; 2001US-280143P.  
 XX (UNIO ) UCB SA.  
 XX Nocka K, Pirozzi G, Einstein R;  
 XX WPI; 2002-508560/54.  
 DR N-PSDB; ABN81324.  
 XX Novel isolated nucleic acids that are differentially expressed in mast  
 PT cells in patients with allergic hypersensitivity, encoding proteins  
 PT associated with mast cell regranulation and allergic hypersensitivity  
 PT -  
 PS Claim 31; Page 117-119; 119pp; English.  
 XX The invention relates to isolated nucleic acid (ABN81319-ABN81324),  
 CC corresponding to genes differentially expressed in mast cells following  
 CC activation or in patients with allergic hypersensitivity disease, (I)  
 CC that encodes proteins (ABN77569-ABN77575) (II) or a protein fragment of  
 CC (II) if at least 6 amino acids. (II) is useful for identifying binding  
 CC partners. (I) or (II) is useful for diagnosing or treating a disease  
 CC state (e.g. allergic hypersensitivity, seasonal rhinitis, asthma,  
 CC urticaria or atopic dermatitis or mastocytosis) in a subject which  
 CC involves determining the level of expression of (I) or (II). A computer  
 CC system, comprising a database containing information identifying the  
 CC expression level in a tissue or at least one mast cell of (I), is useful

CC for presenting information to identify the relative expression level of  
CC (I). (II) is used as a marker to detect, diagnose or identify an allergic  
CC response in a patient. The protein can also serve as a target that  
CC modulate gene expression or activity and as an antigen to raise  
CC polyclonal or monoclonal antibodies. (II) is useful for identifying  
CC agents that modulate expression of the protein or agents, such as  
CC agonists or antagonists. The agonists or antagonists are useful for  
CC modulating biological activity and function of (II) and thus are useful  
CC for alleviating disease conditions such as allergic hypersensitivity,  
CC seasonal rhinitis, asthma, urticaria, atopic dermatitis or mastocytosis.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 23; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQLWVSLRLRHRKAQLLLVNLITFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60  
DB 1 MVQLWVSLRLRHRKAQLLLVNLITFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60

QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVCLDFCGOVCFPLEALLSDLFDPDHCRCQAYSVYAFWISLGGCLGYLLPA 180  
DB 121 ELALLILGVCLDFCGOVCFPLEALLSDLFDPDHCRCQAYSVYAFWISLGGCLGYLLPA 180

QY 181 IDWDTALAPYLGTQECFLGILLTLFTCTVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
DB 181 IDWDTALAPYLGTQECFLGILLTLFTCTVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240

QY 241 CCPCRLARAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFDYDFVGEGL 300  
DB 241 CCPCRLARAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFDYDFVGEGL 300

QY 301 YQGVPAEPCTEARRHYDEGVNRSGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360  
DB 301 YQGVPAEPCTEARRHYDEGVNRSGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

QY 361 APVAAGATCLSHVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 APVAAGATCLSHVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

QY 421 ASSEDSLMTSFLPKPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEA 480  
DB 421 ASSEDSLMTSFLPKPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFNGSIVQLSQSVTAYMWSAAGLGLVAYIFATQ 540  
DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFNGSIVQLSQSVTAYMWSAAGLGLVAYIFATQ 540

QY 541 VFDKSDLAKYSA 553  
DB 541 VFDKSDLAKYSA 553

RESULT 14  
ID ABG61900  
XX ABG61900 standard; Protein; 553 AA.  
AC ABG61900;  
XX  
DT 15-AUG-2002 (first entry)  
DE Prostate cancer-associated protein #101.  
XX Prostate cancer; prostate tumour tissue; human; mammal; cytotostatic.  
KW Mammalia.  
OS  
XX  
PN WC200230268-A2.

XX 18-APR-2002.  
PD  
XX 12-OCT-2001; 2001WO-US32045.  
PF  
XX 13-OCT-2000; 2000US-0687576.  
PR 08-DEC-2000; 2000US-0733288.  
PR 08-DEC-2000; 2000US-0733742.  
PR 24-JAN-2001; 2001US-263957P.  
PR 16-MAR-2001; 2001US-276791P.  
PR 06-APR-2001; 2001US-281922P.  
PR 24-APR-2001; 2001US-286214P.  
PR 30-APR-2001; 2001US-0847046.  
PR 04-MAY-2001; 2001US-288589P.  
XX (EOSB-) EOS BIOTECHNOLOGY INC.  
PA  
XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;  
PI  
XX WPI: 2002-471335/50.  
XX N-PSDB; ABK92217.  
DR  
XX  
XX  
PT Detecting a prostate cancer-associated transcript in a cell in a  
PT patient, useful for diagnosing prostate cancer (PC) or screening  
PT modulators of PC, by determining if prostate cancer-associated genes  
PT are expressed in a prostate tissue -  
XX  
PS Claim 27; Page 386; 436pp; English.  
XX  
CC The present invention relates to methods of detecting a prostate  
CC cancer-associated transcript in a cell from a patient. The method  
CC comprises contacting a biological sample from the patient with  
CC prostate cancer-associated polynucleotides (designated PC genes) that  
CC selectively hybridise to a sequence that is at least 80% identical  
CC to them. The prostate cancer-associated polynucleotide sequences  
CC are differentially expressed in prostate tumour tissue or in  
CC prostate cancer and are derived from the tissues of various  
CC organisms such as humans or other mammals (e.g. mice, sheep and dogs).  
CC The methods of the invention are useful for diagnosing and treating  
CC prostate cancer in mammals. The prostate cancer-associated genes are  
CC useful for diagnosing or treating prostate cancer, as well as for  
CC identifying modulators of prostate cancer or agents that inhibit  
CC prostate cancer. The nucleic acid sequences are particularly useful  
CC in gene therapy, as a vaccine or in antisense applications.  
CC ABG61800-ABG61944 represent prostate cancer-associated proteins.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 23; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQLWVSLRLRHRKAQLLLVNLITFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60  
DB 1 MVQLWVSLRLRHRKAQLLLVNLITFGLVCLAAAGITYVPPLLELVGVBEKFTMWLGIG 60

QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVCLDFCGOVCFPLEALLSDLFDPDHCRCQAYSVYAFWISLGGCLGYLLPA 180  
DB 121 ELALLILGVCLDFCGOVCFPLEALLSDLFDPDHCRCQAYSVYAFWISLGGCLGYLLPA 180

QY 181 IDWDTALAPYLGTQECFLGILLTLFTCTVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240  
DB 181 IDWDTALAPYLGTQECFLGILLTLFTCTVAATLLVAEEAALGPTPEAGLSAPLSLSPH 240

QY 241 CCPCRLARAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFDYDFVGEGL 300  
DB 241 CCPCRLARAFRNIGALLPRHLQCCRMPTLRRLFVAELCSWMLMTFTLFDYDFVGEGL 300

QY 301 YQGVPRAPTEARRHYDEGRMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
Db 301 YQGVPRAPTEARRHYDEGRMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
QY 361 AFPVAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 AFPVAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPPPNGHVCGAGSGLLPPPPALCGASACDVSVRVVVGPTPEA 480  
Db 421 ASSEDSLMTSFLPGPKGAPPPNGHVCGAGSGLLPPPPALCGASACDVSVRVVVGPTPEA 480  
QY 481 RVVPGRGICLDLAILDSAPFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVATQ 540  
Db 481 RVVPGRGICLDLAILDSAPFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

RESULT 15

ABB95222 standard; Protein; 553 AA.

AC ABB95222;

DT 19-JUL-2002 (first entry)

XX Human L1-12 protein SEQ ID NO 113.

DE Human; cancer; prostate cancer; vaccine; cytostatic; immunostimulant;  
KW gene therapy.

OS Homo sapiens.

PN US2002022248-A1.

XX 21-FEB-2002.

PF 12-JAN-2001; \*2001US-0759143.

XX 25-FEB-1997; 97US-0806099.  
PR 01-AUG-1997; 97US-0904804.  
PR 09-FEB-1998; 98US-020956.  
PR 25-FEB-1998; 98US-0030607.  
PR 14-JUL-1998; 98US-0030607.  
PR 23-SEP-1998; 98US-0159812.  
PR 15-JAN-1999; 99US-0232149.  
PR 09-APR-1999; 99US-0288946.  
PR 13-JUL-1999; 99US-0352616.  
PR 12-NOV-1999; 99US-0439313.  
PR 18-NOV-1999; 99US-0443686.  
PR 14-JAN-2000; 2000US-0483672.  
PR 27-MAR-2000; 2000US-0536857.  
PR 09-MAY-2000; 2000US-0568100.  
PR 12-MAY-2000; 2000US-0570737.  
PR 13-JUN-2000; 2000US-0593793.  
PR 27-JUN-2000; 2000US-0605783.  
PR 10-AUG-2000; 2000US-0636215.  
PR 29-AUG-2000; 2000US-0651236.  
PR 06-SEP-2000; 2000US-0657279.  
PR 02-OCT-2000; 2000US-0679426.  
PR 10-OCT-2000; 2000US-0685166.

XX (XUJJ/) XU J.  
PA (DILL/) DILLON D C.  
PA (MITC/) MITCHAM J L.  
PA (HARL/) HARLOCKER S L.  
PA (JIAN/) JIANG Y.  
PA (KALO/) KALOS M D.  
PA (FANG/) FANGER G R.  
PA (RETT/) RETTER M W.

PA (STOL/) STOLK J A.  
PA (DAYC/) DAY C H.  
PA (VEDV/) VEDVICK T S.  
PA (CART/) CARTER D.  
PA (LISX/) LI S X.  
PA (WANG/) WANG A.  
PA (SKEI/) SKEIKY Y A W.  
PA (HEPL/) HEPLER W T.  
PA (HEND/) HENDERSON R A.  
XX  
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;  
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;  
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;  
XX  
DR WPI; 2002-255649/30.  
XX  
PT New prostate-specific polynucleotides for diagnosing and treating  
PT diseases, in particular prostate cancer, and as markers for the  
PT progression of cancer  
XX  
PS Claim 2; SEQ ID NO 113; 87pp; English.  
XX  
CC The present invention provides prostate-specific coding sequences and  
CC their encoded proteins. These can be used in the diagnosis and treatment  
CC of cancers, particularly prostate cancer. The present sequence is a  
CC protein described in the invention.  
XX  
SQ Sequence 553 AA;

Query Match 100.0%; Score 553; DB 23; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHKAQILLNLLTFGLVCLAAAGTIVPPILLLEVGVBEKMTWVLGIG 60  
Db 1 MVQRLWVSRLLRHKAQILLNLLTFGLVCLAAAGTIVPPILLLEVGVBEKMTWVLGIG 60  
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
Db 61 PVLGLVCVPLLGASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGQVCFTPLEALLSDLPDPDHCQAYSVYAFMISLGGCLGYLLPA 180  
Db 121 ELALLILGVGLDFCGQVCFTPLEALLSDLPDPDHCQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTSAAPYLGTOEBCIFGLLLTLITCTVAATLLVAEEAALGPTBPAGLSAPLSPH 240  
Db 181 IDWDTSAAPYLGTOEBCIFGLLLTLITCTVAATLLVAEEAALGPTBPAGLSAPLSPH 240  
QY 241 CCPCARLAFRNIGALLPRLHQLCCRMPTLRLFVAELCSMMALMTPTLFYTDVFGEL 300  
Db 241 CCPCARLAFRNIGALLPRLHQLCCRMPTLRLFVAELCSMMALMTPTLFYTDVFGEL 300  
QY 301 YQGVPRAPTEARRHYDEGRMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
Db 301 YQGVPRAPTEARRHYDEGRMGSLGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360  
QY 361 AFPVAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
Db 361 AFPVAGATCLSHSVAVVTASAAITGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPPPNGHVCGAGSGLLPPPPALCGASACDVSVRVVVGPTPEA 480  
Db 421 ASSEDSLMTSFLPGPKGAPPPNGHVCGAGSGLLPPPPALCGASACDVSVRVVVGPTPEA 480  
QY 481 RVVPGRGICLDLAILDSAPFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVATQ 540  
Db 481 RVVPGRGICLDLAILDSAPFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVATQ 540  
QY 541 VVFDKSDLAKYSA 553  
Db 541 VVFDKSDLAKYSA 553

Search completed: December 3, 2003, 17:23:03  
Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:22:07 ; Search time 36 seconds  
(without alignments)  
2856.916 Million cell updates/sec

Title: US-09-593-793A-113

Perfect score: 553

Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFVDSIAKYSA 553

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 684280 seqs, 185983659 residues

Word size : 0

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- Published Applications AA:\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
  - 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
  - 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
  - 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
  - 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
  - 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
  - 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	553	100.0	553	9	US-09-745-288-101
2	553	100.0	553	9	US-09-838-785-2
3	553	100.0	553	9	US-09-759-143-113
4	553	100.0	553	9	US-09-780-669-113
5	553	100.0	553	9	US-09-030-606-113
6	553	100.0	553	9	US-09-822-827-113
7	553	100.0	553	9	US-09-115-453-113
8	553	100.0	553	10	US-09-232-880-113
9	553	100.0	553	10	US-09-895-793-113
10	553	100.0	553	10	US-09-895-814-113
11	553	100.0	553	12	US-10-144-678A-113
12	553	100.0	553	12	US-10-005-907-113
13	553	100.0	553	12	US-10-294-025-113
14	553	100.0	553	14	US-10-012-896-113
15	553	100.0	553	15	US-10-010-940-113

16	553	100.0	1079	9	US-09-822-827-947	Sequence 947, App
17	553	100.0	1079	10	US-09-895-793-947	Sequence 947, App
18	501	90.6	710	15	US-10-296-770-4	Sequence 4, Appli
19	320	57.9	359	9	US-09-822-827-974	Sequence 974, App
20	320	57.9	359	10	US-09-895-793-974	Sequence 974, App
21	314	56.8	530	15	US-10-296-770-5	Sequence 5, Appli
22	297	53.7	305	12	US-10-144-678A-1029	Sequence 1029, Ap
23	297	53.7	305	12	US-10-294-025-1029	Sequence 1029, Ap
24	263	47.6	371	9	US-09-759-143-708	Sequence 708, App
25	263	47.6	371	9	US-09-780-669-708	Sequence 708, App
26	263	47.6	371	9	US-09-822-827-708	Sequence 708, App
27	263	47.6	371	10	US-09-895-793-708	Sequence 708, App
28	263	47.6	371	10	US-09-895-814-708	Sequence 708, App
29	263	47.6	371	12	US-10-144-678A-708	Sequence 708, App
30	263	47.6	371	12	US-10-294-025-708	Sequence 708, App
31	263	47.6	371	14	US-10-012-896-708	Sequence 708, App
32	263	47.6	400	9	US-09-759-143-852	Sequence 852, App
33	263	47.6	400	9	US-09-780-669-852	Sequence 852, App
34	263	47.6	400	9	US-09-822-827-852	Sequence 852, App
35	263	47.6	400	10	US-09-895-793-852	Sequence 852, App
36	263	47.6	400	10	US-09-895-814-852	Sequence 852, App
37	263	47.6	400	12	US-10-144-678A-852	Sequence 852, App
38	263	47.6	400	12	US-10-294-025-852	Sequence 852, App
39	263	47.6	400	14	US-10-012-896-852	Sequence 852, App
40	238	43.0	246	12	US-10-144-678A-1028	Sequence 1028, Ap
41	238	43.0	246	12	US-10-294-025-1028	Sequence 1028, Ap
42	187	33.8	326	12	US-10-005-907-5	Sequence 5, Appli
43	187	33.8	355	12	US-10-144-678A-1011	Sequence 1011, Ap
44	187	33.8	355	12	US-10-294-025-1011	Sequence 1011, Ap
45	187	33.8	355	14	US-10-012-896-1011	Sequence 1011, Ap

ALIGNMENTS

RESULT 1

US-09-745-288-101  
; Sequence 101, Application US/09745288  
; Patent No. US20010018058A1  
; GENERAL INFORMATION:  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE  
; FILE REFERENCE: 210121.446D1  
; CURRENT APPLICATION NUMBER: US/09/745,288  
; CURRENT FILING DATE: 2000-12-19  
; NUMBER OF SEQ ID NOS: 101  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 101  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-745-288-101

Query Match	100.0%	Score 553;	DB 9;	Length 553;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 553;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	MYQRLWVSRLLRHRKAQLLLVNLITFGLEVCLAAAGITVVPPLLLLEVGVVEEKFTWVLGIG	60	
Db	1	MYQRLWVSRLLRHRKAQLLLVNLITFGLEVCLAAAGITVVPPLLLLEVGVVEEKFTWVLGIG	60	
Qy	61	PVLGLVCPVLLGSASDHWRGRRPFTWALSGLLSLFLIPRAGMLAGLCCPDRL	120	
Db	61	PVLGLVCPVLLGSASDHWRGRRPFTWALSGLLSLFLIPRAGMLAGLCCPDRL	120	
Qy	121	ELALLILVGLLDFCGQVCFPLEALLSDLPDPHCRQAVSYAFMISLGCIGYLLPA	180	
Db	121	ELALLILVGLLDFCGQVCFPLEALLSDLPDPHCRQAVSYAFMISLGCIGYLLPA	180	
Qy	181	IDWDTLSALAPVLTQEECLFGLLTILFTCTVAATLLVAERAAALGPTPEAGLSATSLSPH	240	

```

Db      181  IDWDTSAALPYLGTQBECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240
Qy      241  CCPCARLAFRNLGALLPRHLQCCMPRTLRRLFVAELCSMMALMTFTLFYDFVGBGL 300
Db      241  CCPCARLAFRNLGALLPRHLQCCMPRTLRRLFVAELCSMMALMTFTLFYDFVGBGL 300
Qy      301  YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
Db      301  YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
Qy      361  APFVAAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db      361  APFVAAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Qy      421  ASSEDSMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480
Db      421  ASSEDSMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480
Qy      481  RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
Db      481  RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
Qy      541  VVFDKSDLAKYSA 553
Db      541  VVFDKSDLAKYSA 553

```

RESULT 2

```

US-09-838-785-2
; Sequence 2, Application US/09838785
; Patent No. US2002009455A1
; GENERAL INFORMATION:
; APPLICANT: Lau, Ted
; APPLICANT: Lin, Rick
; APPLICANT: Parkes, Debbie
; APPLICANT: Party, Gordon
; APPLICANT: Schneider, Douglas
; APPLICANT: Steinbrecher, Renate
; APPLICANT: Van Heuit, Pam T
; APPLICANT: Wu, John
; TITLE OF INVENTION: DNA Encoding a No. US2002009455A1el PROST Q3
; FILE REFERENCE: 51831AUSM1
; CURRENT APPLICATION NUMBER: US/09/838,785
; CURRENT FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: 60/200,065
; PRIOR FILING DATE: 2000-04-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-838-785-2

```

```

Query Match      100.0%; Score 553; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLLEVGVEEKFMTWLIGIG 60
Db      1  MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLLEVGVEEKFMTWLIGIG 60
Qy      61  PVGLVVCVPLLGASDHWGRGRRRPPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
Db      61  PVGLVVCVPLLGASDHWGRGRRRPPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
Qy      121  ELALLILVGLDFFCCQVCFPLEALLSFLFRPDHCRQAYSVYAFWISIGGCLGYLLPA 180
Db      121  ELALLILVGLDFFCCQVCFPLEALLSFLFRPDHCRQAYSVYAFWISIGGCLGYLLPA 180
Qy      181  IDWDTSAALPYLGTQBECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240

```

```

Db      181  IDWDTSAALPYLGTQBECLFGLLTLLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240
Qy      241  CCPCARLAFRNLGALLPRHLQCCMPRTLRRLFVAELCSMMALMTFTLFYDFVGBGL 300
Db      241  CCPCARLAFRNLGALLPRHLQCCMPRTLRRLFVAELCSMMALMTFTLFYDFVGBGL 300
Qy      301  YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
Db      301  YQGVPAEPTGTEARRHYDEGVRMGSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
Qy      361  APFVAAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Db      361  APFVAAGATCLSHSVAVVTASAALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
Qy      421  ASSEDSMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480
Db      421  ASSEDSMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480
Qy      481  RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
Db      481  RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
Qy      541  VVFDKSDLAKYSA 553
Db      541  VVFDKSDLAKYSA 553

```

RESULT 3

```

US-09-759-143-113
; Sequence 113, Application US/09759143
; Patent No. US200202248A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqui
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C23
; CURRENT APPLICATION NUMBER: US/09/759,143
; CURRENT FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-759-143-113

```

```

Query Match      100.0%; Score 553; DB 9; Length 553;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLLEVGVEEKFMTWLIGIG 60
Db      1  MVQRLVWSRLRRKRAQLLLVNLTFGLVCLAAAGITYVPPLLLEVGVEEKFMTWLIGIG 60
Qy      61  PVGLVVCVPLLGASDHWGRGRRRPPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120
Db      61  PVGLVVCVPLLGASDHWGRGRRRPPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120

```

Qy	121	ELALLILGVLLDFCGVCF	TPLEALLS	DLFRDP	PHCRQAYSVYAF	PMISLG	GGCLGYLLPA	180			
Db	121	ELALLILGVLLDFCGVCF	TPLEALLS	DLFRDP	PHCRQAYSVYAF	PMISLG	GGCLGYLLPA	180			
Qy	181	IDWDTSA	LAPYLGTQ	EECLFGLL	TLTILFTCV	AAATLLVA	EEAALGPTEPA	EGLSAPSLSPH 240			
Db	181	IDWDTSA	LAPYLGTQ	EECLFGLL	TLTILFTCV	AAATLLVA	EEAALGPTEPA	EGLSAPSLSPH 240			
Qy	241	CCPCAR	LA	AF	RNLG	ALLP	RLHOLCC	RMPTLRLFVAELCS	NMALMTFTL	FYDFVGEGL 300	
Db	241	CCPCAR	LA	AF	RNLG	ALLP	RLHOLCC	RMPTLRLFVAELCS	NMALMTFTL	FYDFVGEGL 300	
Qy	301	YQGVPR	ABPGTE	ARRHYDEG	VRMGSLGL	FLQCAISL	VFLV	NMDRLV	QRFEGTRAV	YLASVA 360	
Db	301	YQGVPR	ABPGTE	ARRHYDEG	VRMGSLGL	FLQCAISL	VFLV	NMDRLV	QRFEGTRAV	YLASVA 360	
Qy	361	APFVAAG	ATCLSHSVAV	VTASAA	LGTFTS	ALQIILPY	TLASIV	YHREK	OVFLPKY	RGDTCG 420	
Db	361	APFVAAG	ATCLSHSVAV	VTASAA	LGTFTS	ALQIILPY	TLASIV	YHREK	OVFLPKY	RGDTCG 420	
Qy	421	ASSBDS	SLMTS	FLPGPK	GAPFPNG	HVGAGSG	LLPPP	ALCGAS	CDVS	SVRVVVGEPTEA 480	
Db	421	ASSBDS	SLMTS	FLPGPK	GAPFPNG	HVGAGSG	LLPPP	ALCGAS	CDVS	SVRVVVGEPTEA 480	
Qy	481	RVVPGRG	ICLD	LA	ILDS	AF	LLSQVAPS	IFMGSIV	QLSQSV	TVMYWSAAGLGH	VALYFATQ 540
Db	481	RVVPGRG	ICLD	LA	ILDS	AF	LLSQVAPS	IFMGSIV	QLSQSV	TVMYWSAAGLGH	VALYFATQ 540
Qy	541	VVFDKSD	LAKYSA	553							
Db	541	VVFDKSD	LAKYSA	553							

## RESULT 4

US-09-780-669-113 -  
; Sequence 113, Application US/09780669  
; Patent No. US20020051977A1

APPLICANT:	Xu, Jiangchun
APPLICANT:	Dillon, Davin C.
APPLICANT:	Mitcham, Jennifer L.
APPLICANT:	Harlocker, Susan L.
APPLICANT:	Jiang, Yuqi
APPLICANT:	Henderson, Robert A.
APPLICANT:	Kalos, Michael D.
APPLICANT:	Fanger, Gary R.
APPLICANT:	Rester, Marc W.
APPLICANT:	Stolk, John A.
APPLICANT:	Day, Craig H.
APPLICANT:	Vedvick, Thomas S.
APPLICANT:	Carter, Patrick
APPLICANT:	Li, Samuel
APPLICANT:	Wang, Aijun
APPLICANT:	Skeiky, Yasir A.W.
APPLICANT:	Hepler, William
APPLICANT:	Hural, John
APPLICANT:	McNeill, Patricia D.
APPLICANT:	Houghton, Raymond L.

; AFFILIATION: Houghton, Raymond L.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
; TITLE OF INVENTION:

```

: FILE REFERENCE: 210121.427C24
: CURRENT APPLICATION NUMBER: US/09/780.669
: CURRENT FILING DATE: 2001-02-09
: NUMBER OF SEQ ID NOS: 943
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 113
: LENGTH: 553

```

; ORGANISM: Homo  
US-09-780-669-113

Query Match 100.0%; Score 553; DB 9; Length 553;

[illegible]

## RESULT. T 5

US-09-030-606-113  
; Sequence 113, Application US/09030606  
; Patent No. US20020081580A1

```

; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS F
; NUMBER OF SEQUENCES: 224

```

22.  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104

```

? ZIP: 98104
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/030,606
? FILING DATE: 25-FEB-1998

```

; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Maki, David J.  
 ; REGISTRATION NUMBER: 31,392  
 ; REFERENCE/DOCKET NUMBER: 210121.428C3

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 113:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 553 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 US-09-030-606-113

Query Match 100.0%; Score 553; DB 9; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQQLLVNLLTFGLVCLAAAGITYVPPLLLLEVGVVEEKFTMWLGIG 60  
 DB 1 MVQRLWVSRLLRHRKAQQLLVNLLTFGLVCLAAAGITYVPPLLLLEVGVVEEKFTMWLGIG 60

QY 61 PVLGLVCVPLGASDHWGRYGRRRPFIWALSIGLILSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVLGLVCVPLGASDHWGRYGRRRPFIWALSIGLILSLFLIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVABEAAALGPTPEAGLSAPLSHPH 240  
 DB 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVABEAAALGPTPEAGLSAPLSHPH 240

QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300  
 DB 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300

QY 301 YQGVPAEPGTEARRHYDEGRVMSIGLFLQCAISLVFSLVMDRLVQRFEGTRAVYLASVA 360  
 DB 301 YQGVPAEPGTEARRHYDEGRVMSIGLFLQCAISLVFSLVMDRLVQRFEGTRAVYLASVA 360

QY 361 APVPAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 APVPAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
 DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAAGLGLVAYIFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAAGLGLVAYIFATQ 540

QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

## RESULT 6

US-09-822-827-113  
 Sequence 113, Application US/09822827  
 Patent No. US20020081680A1  
 GENERAL INFORMATION:  
 APPLICANT: Xu, Jiangchun  
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 FILE REFERENCE: 210121.534C1  
 CURRENT APPLICATION NUMBER: US/09/822,827  
 NUMBER OF SEQ ID NOS: 982  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 113

LENGTH: 553  
 TYPE: PRT  
 ORGANISM: Homo sapien  
 US-09-822-827-113

Query Match 100.0%; Score 553; DB 9; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQQLLVNLLTFGLVCLAAAGITYVPPLLLLEVGVVEEKFTMWLGIG 60  
 DB 1 MVQRLWVSRLLRHRKAQQLLVNLLTFGLVCLAAAGITYVPPLLLLEVGVVEEKFTMWLGIG 60

QY 61 PVLGLVCVPLGASDHWGRYGRRRPFIWALSIGLILSLFLIPRAGWLAGLLCPDPRPL 120  
 DB 61 PVLGLVCVPLGASDHWGRYGRRRPFIWALSIGLILSLFLIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180  
 DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCRCQAYSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVABEAAALGPTPEAGLSAPLSHPH 240  
 DB 181 IDWDTSAALPYLGTQECLFGLLTFLITCVAATLLVABEAAALGPTPEAGLSAPLSHPH 240

QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300  
 DB 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300

QY 301 YQGVPAEPGTEARRHYDEGRVMSIGLFLQCAISLVFSLVMDRLVQRFEGTRAVYLASVA 360  
 DB 301 YQGVPAEPGTEARRHYDEGRVMSIGLFLQCAISLVFSLVMDRLVQRFEGTRAVYLASVA 360

QY 361 APVPAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420  
 DB 361 APVPAAGATCLSHSVAVVTASALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480  
 DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAAGLGLVAYIFATQ 540  
 DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAAGLGLVAYIFATQ 540

QY 541 VVFDKSDLAKYSA 553  
 DB 541 VVFDKSDLAKYSA 553

## RESULT 7

US-09-115-453-113  
 Sequence 113, Application US/09115453B  
 Patent No. US20020090372A1  
 GENERAL INFORMATION:  
 APPLICANT: Xu, Jiangchun  
 APPLICANT: Dillon, Davin C.  
 TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND  
 FILE REFERENCE: 210121.427C4  
 CURRENT APPLICATION NUMBER: US/09/115,453B  
 CURRENT FILING DATE: 1998-07-14  
 NUMBER OF SEQ ID NOS: 228  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 113  
 LENGTH: 553  
 TYPE: PRT  
 ORGANISM: Homo sapien  
 US-09-115-453-113

Query Match 100.0%; Score 553; DB 9; Length 553;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```
QY 1 MVORLWVSRLLRRHKAQILLVLLTFLGLEVCLAAAGITYVPPLLLLEGVGVEEKFTMWLIG 60
DB 1 MVORLWVSRLLRRHKAQILLVLLTFLGLEVCLAAAGITYVPPLLLLEGVGVEEKFTMWLIG 60
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
QY 121 ELALLIIGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
DB 121 ELALLIIGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSALAPYLGTQBECLFGLLTFLITCVAAATLLVAEEAALGPTPEAEGLSAPLSPH 240
DB 181 IDWDTSALAPYLGTQBECLFGLLTFLITCVAAATLLVAEEAALGPTPEAEGLSAPLSPH 240
QY 241 CCPCRARLAFRNIGALLPRHOLCCRMPTLRRLFVAELCSMWALMTFTLYTDFVGEGL 300
DB 241 CCPCRARLAFRNIGALLPRHOLCCRMPTLRRLFVAELCSMWALMTFTLYTDFVGEGL 300
QY 301 YQGVPRABEPTGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
DB 301 YQGVPRABEPTGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
QY 481 RVVPRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAIYVYFATQ 540
DB 481 RVVPRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAIYVYFATQ 540
QY 541 VVFDKSDLAKYSA 553
DB 541 VVFDKSDLAKYSA 553

RESULT 8
US-09-232-880-113
; Sequence 113, Application US/09232880
; Publication No. US20020182596A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF
; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.428C6
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-232-880-113

Query Match 100.0%; Score 553; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVORLWVSRLLRRHKAQILLVLLTFLGLEVCLAAAGITYVPPLLLLEGVGVEEKFTMWLIG 60
DB 1 MVORLWVSRLLRRHKAQILLVLLTFLGLEVCLAAAGITYVPPLLLLEGVGVEEKFTMWLIG 60
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
QY 121 ELALLIIGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
DB 121 ELALLIIGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSALAPYLGTQBECLFGLLTFLITCVAAATLLVAEEAALGPTPEAEGLSAPLSPH 240
DB 181 IDWDTSALAPYLGTQBECLFGLLTFLITCVAAATLLVAEEAALGPTPEAEGLSAPLSPH 240
QY 241 CCPCRARLAFRNIGALLPRHOLCCRMPTLRRLFVAELCSMWALMTFTLYTDFVGEGL 300
DB 241 CCPCRARLAFRNIGALLPRHOLCCRMPTLRRLFVAELCSMWALMTFTLYTDFVGEGL 300
QY 301 YQGVPRABEPTGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
DB 301 YQGVPRABEPTGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
QY 481 RVVPRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAIYVYFATQ 540
DB 481 RVVPRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAIYVYFATQ 540
QY 541 VVFDKSDLAKYSA 553
DB 541 VVFDKSDLAKYSA 553

RESULT 9
US-09-895-793-113
; Sequence 113, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuxiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darriek
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
```

```
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFIWALSIGILLSLFLIPRAGWLAGLCPDPRPL 120
QY 121 ELALLIIGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
DB 121 ELALLIIGVGLDFCGQVCFPLEALLSDFRDPDHCROAYSVYAFMISLGGCLGYLLPA 180
QY 181 IDWDTSALAPYLGTQBECLFGLLTFLITCVAAATLLVAEEAALGPTPEAEGLSAPLSPH 240
DB 181 IDWDTSALAPYLGTQBECLFGLLTFLITCVAAATLLVAEEAALGPTPEAEGLSAPLSPH 240
QY 241 CCPCRARLAFRNIGALLPRHOLCCRMPTLRRLFVAELCSMWALMTFTLYTDFVGEGL 300
DB 241 CCPCRARLAFRNIGALLPRHOLCCRMPTLRRLFVAELCSMWALMTFTLYTDFVGEGL 300
QY 301 YQGVPRABEPTGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
DB 301 YQGVPRABEPTGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQREGTRAVYLASVA 360
QY 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
DB 361 AFPVAAGATCLSHSVAVVTASAAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
QY 481 RVVPRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAIYVYFATQ 540
DB 481 RVVPRGICLDLAILDSAPLLSQVAPSLFMGSIQVLSQVTAIYVYFATQ 540
QY 541 VVFDKSDLAKYSA 553
DB 541 VVFDKSDLAKYSA 553

RESULT 9
US-09-895-793-113
; Sequence 113, Application US/09895793
; Publication No. US20020192763A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuxiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darriek
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.534C2
; CURRENT APPLICATION NUMBER: US/09/895,793
; CURRENT FILING DATE: 2001-06-29
; NUMBER OF SEQ ID NOS: 982
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
```

US-09-895-793-113

```
Query Match      100.0%; Score 553; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQALLVNLITFGLVCLAAAGITYVPPILLLEVGVEEKFTMWLGIG 60
   |||||
Db 1 MVQRLWVSRLLRHRKAQALLVNLITFGLVCLAAAGITYVPPILLLEVGVEEKFTMWLGIG 60
   |||||

QY 61 PVLGLVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGMLAGLLCPDPRPL 120
   |||||
Db 61 PVLGLVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGMLAGLLCPDPRPL 120
   |||||

QY 121 ELALLITLGVLLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGCLGYLLPA 180
   |||||
Db 121 ELALLITLGVLLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGCLGYLLPA 180
   |||||

QY 181 IDWDTLSALAPYLGTQECFLGILLTLIFLTCVAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240
   |||||
Db 181 IDWDTLSALAPYLGTQECFLGILLTLIFLTCVAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240
   |||||

QY 241 CCPCRLARLAFNRLGALLPRLHQLCCMRPTRLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300
   |||||
Db 241 CCPCRLARLAFNRLGALLPRLHQLCCMRPTRLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300
   |||||

QY 301 YQGVPRAPGTEARRHYDEGVGMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLASVA 360
   |||||
Db 301 YQGVPRAPGTEARRHYDEGVGMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLASVA 360
   |||||

QY 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
   |||||
Db 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
   |||||

QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
   |||||
Db 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
   |||||

QY 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIVOLQSQVTAVMVSAAGLGLVAYIFATQ 540
   |||||
Db 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIVOLQSQVTAVMVSAAGLGLVAYIFATQ 540
   |||||

QY 541 VVFDKSDLAKYSA 553
   |||||
Db 541 VVFDKSDLAKYSA 553
   |||||
```

RESULT 10

```
US-09-895-814-113
; Sequence 113, Application US/09895814
; Publication No. US20020193296A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Xuqiu
; APPLICANT: Kalos, Michael D.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darriick
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Hural, John
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Vinals de Bassols, Carlota
; APPLICANT: Foy, Teresa
```

```
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C26
; CURRENT APPLICATION NUMBER: US/09/895,814
; NUMBER OF SEQ ID NOS: 990
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 113
; LENGTH: 553
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-895-814-113
```

```
Query Match      100.0%; Score 553; DB 10; Length 553;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVQRLWVSRLLRHRKAQALLVNLITFGLVCLAAAGITYVPPILLLEVGVEEKFTMWLGIG 60
   |||||
Db 1 MVQRLWVSRLLRHRKAQALLVNLITFGLVCLAAAGITYVPPILLLEVGVEEKFTMWLGIG 60
   |||||

QY 61 PVLGLVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGMLAGLLCPDPRPL 120
   |||||
Db 61 PVLGLVCVPLLGASADHWRGRRRPFIFWALSGLILLSLFLIPRAGMLAGLLCPDPRPL 120
   |||||

QY 121 ELALLITLGVLLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGCLGYLLPA 180
   |||||
Db 121 ELALLITLGVLLDFCGQVCFPLEALLSDLPDPDHCROAYSVYAFMISLGCLGYLLPA 180
   |||||

QY 181 IDWDTLSALAPYLGTQECFLGILLTLIFLTCVAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240
   |||||
Db 181 IDWDTLSALAPYLGTQECFLGILLTLIFLTCVAATLLVAEEAALGPTPEAEGLSAPLSLSPH 240
   |||||

QY 241 CCPCRLARLAFNRLGALLPRLHQLCCMRPTRLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300
   |||||
Db 241 CCPCRLARLAFNRLGALLPRLHQLCCMRPTRLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300
   |||||

QY 301 YQGVPRAPGTEARRHYDEGVGMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLASVA 360
   |||||
Db 301 YQGVPRAPGTEARRHYDEGVGMGSLGLFLOCAISLVFSLVMDRLVQVREGTRAVYLASVA 360
   |||||

QY 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
   |||||
Db 361 APFVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
   |||||

QY 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
   |||||
Db 421 ASSEDSLMTSFLPGPKPGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
   |||||

QY 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIVOLQSQVTAVMVSAAGLGLVAYIFATQ 540
   |||||
Db 481 RVVPGRGICLDLAILDLSAFLLSQVAPSLFMGSIVOLQSQVTAVMVSAAGLGLVAYIFATQ 540
   |||||

QY 541 VVFDKSDLAKYSA 553
   |||||
Db 541 VVFDKSDLAKYSA 553
   |||||
```

RESULT 11

```
US-10-144-678A-113
; Sequence 113, Application US/10144678A
; Publication No. US20030157089A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Xuqiu
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
```

APPLICANT: Stolk, John A.  
APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A. W.  
APPLICANT: Hepler, William T.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals y de Basols, Carlota  
APPLICANT: Foy, Teresa M.  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Deng, Ta  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.427C28  
CURRENT APPLICATION NUMBER: US/10/144, 678A  
CURRENT FILING DATE: 2002-08-12  
NUMBER OF SEQ ID NOS: 1033  
SOFTWARE: Fast-Seq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-144-678A-113

Query Match 100.0%; Score 553; DB 12; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVORLWVSRLRRKKAQLLAVNLLTFTGLEVCLAAGTYVPPLLLVGVVEKEKMTWVLGIG 60  
DB 1 MVORLWVSRLRRKKAQLLAVNLLTFTGLEVCLAAGTYVPPLLLVGVVEKEKMTWVLGIG 60  
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCROQAYSVYAFMISLGGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCROQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTLSALAPYLGTOECLFGLLTFLITCVAATLLVABEALGPTPEAEGLSAPLSHP 240  
DB 181 IDWDTLSALAPYLGTOECLFGLLTFLITCVAATLLVABEALGPTPEAEGLSAPLSHP 240  
QY 241 CCPCARLAFRNIGALLPRLHQLCCRMPTLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
DB 241 CCPCARLAFRNIGALLPRLHQLCCRMPTLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
QY 301 YQGVPRABPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
DB 301 YQGVPRABPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
QY 361 APVVAAGATCLSHSVAVVTASALTGTFTSALOILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 APVVAAGATCLSHSVAVVTASALTGTFTSALOILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMYSAAGLGLVAIFYFATQ 540  
DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMYSAAGLGLVAIFYFATQ 540  
QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553

RESULT 13  
US-10-294-025-113  
Sequence 113, Application US/10294025  
Publication No. US20030185830A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun

RESULT 12  
US-10-005-907-13  
Sequence 13, Application US/10005907  
Publication No. US20030166881A1  
GENERAL INFORMATION:  
APPLICANT: Union Chimique Belge, S.A.  
APPLICANT: No. US20030166881A1ka, Karl  
APPLICANT: Pirozzi, Gregory  
APPLICANT: Einstein, Richard  
TITLE OF INVENTION: NOVEL GENES ASSOCIATED WITH ALLERGIC HYPERSENSITIVITY AND MAST CELL  
TITLE OF INVENTION: ACTIVATION  
FILE REFERENCE: 053529-5005  
CURRENT APPLICATION NUMBER: US/10/005,907  
CURRENT FILING DATE: 2001-12-07  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 13  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-005-907-13

Query Match 100.0%; Score 553; DB 12; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MVORLWVSRLRRKKAQLLAVNLLTFTGLEVCLAAGTYVPPLLLVGVVEKEKMTWVLGIG 60  
DB 1 MVORLWVSRLRRKKAQLLAVNLLTFTGLEVCLAAGTYVPPLLLVGVVEKEKMTWVLGIG 60  
QY 61 PVLGLVCVPLLGASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
DB 61 PVLGLVCVPLLGASDHWGRYGRRRPFTWALSIGILLSLFLIPRAGWLAGLLCPDPRPL 120  
QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCROQAYSVYAFMISLGGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFLRDPDHCROQAYSVYAFMISLGGCLGYLLPA 180  
QY 181 IDWDTLSALAPYLGTOECLFGLLTFLITCVAATLLVABEALGPTPEAEGLSAPLSHP 240  
DB 181 IDWDTLSALAPYLGTOECLFGLLTFLITCVAATLLVABEALGPTPEAEGLSAPLSHP 240  
QY 241 CCPCARLAFRNIGALLPRLHQLCCRMPTLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
DB 241 CCPCARLAFRNIGALLPRLHQLCCRMPTLRRLFVAELCSMMALMTFTLFTYDFVGEGL 300  
QY 301 YQGVPRABPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
DB 301 YQGVPRABPGTEARRHYDEGVNMGSLGLFLOCAISLVFSLVMDRLVQRFGTTRAVYLASVA 360  
QY 361 APVVAAGATCLSHSVAVVTASALTGTFTSALOILPYTLASLYHREKQVFLPKYRGDTGG 420  
DB 361 APVVAAGATCLSHSVAVVTASALTGTFTSALOILPYTLASLYHREKQVFLPKYRGDTGG 420  
QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPEPTA 480  
QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMYSAAGLGLVAIFYFATQ 540  
DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMYSAAGLGLVAIFYFATQ 540  
QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553

RESULT 13  
US-10-294-025-113  
Sequence 113, Application US/10294025  
Publication No. US20030185830A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun

APPLICANT: Stolk, John A.  
APPLICANT: Kalos, Michael D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C29  
CURRENT APPLICATION NUMBER: US/10/294,025  
CURRENT FILING DATE: 2002-11-12  
NUMBER OF SEQ ID NOS: 1038  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-294-025-113

Query Match 100.0%; Score 553; DB 12; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVORLWVSLRHRKAQILLVNLITFGLVCLAAAGTIVPPILLLEVGVBEKMTWVLGIG 60  
DB 1 MVORLWVSLRHRKAQILLVNLITFGLVCLAAAGTIVPPILLLEVGVBEKMTWVLGIG 60

QY 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSGLILLSLFIIPRAGWLAGLLCPDPRPL 120  
DB 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSGLILLSLFIIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROQVSVYAFMISLGGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROQVSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSALAPYLGTQBECLFGLITLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
DB 181 IDWDTSALAPYLGTQBECLFGLITLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240

QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300  
DB 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300

QY 301 YQGVPAEPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
DB 301 YQGVPAEPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360

QY 361 APPVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTG 420  
DB 361 APPVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTG 420

QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGGLPPPPALCGASACDVSVRVVVGEPTEA 480  
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGGLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAYIFATQ 540  
DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAYIFATQ 540

QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553

## RESULT 14

US-10-012-896-113  
Sequence 113, Application US/10012896  
Publication No. US20020183251A1  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Dillon, Davin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Jiang, Yuqiu  
APPLICANT: Kalos, Michael D.  
APPLICANT: Retter, Marc W.  
APPLICANT: Stolk, John A.

APPLICANT: Day, Craig H.  
APPLICANT: Vedvick, Thomas S.  
APPLICANT: Carter, Darrick  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
APPLICANT: Hural, John  
APPLICANT: McNeill, Patricia D.  
APPLICANT: Houghton, Raymond L.  
APPLICANT: Vinals de Bassols, Carlota  
APPLICANT: Foy, Teresa  
APPLICANT: Fanger, Gary R.  
APPLICANT: Watanabe, Yoshihiro  
APPLICANT: Meagher, Madeleine Joy  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 210121.427C27  
CURRENT APPLICATION NUMBER: US/10/012,896  
CURRENT FILING DATE: 2001-12-10  
NUMBER OF SEQ ID NOS: 1011  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 113  
LENGTH: 553  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-012-896-113

Query Match 100.0%; Score 553; DB 14; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVORLWVSLRHRKAQILLVNLITFGLVCLAAAGTIVPPILLLEVGVBEKMTWVLGIG 60  
DB 1 MVORLWVSLRHRKAQILLVNLITFGLVCLAAAGTIVPPILLLEVGVBEKMTWVLGIG 60

QY 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSGLILLSLFIIPRAGWLAGLLCPDPRPL 120  
DB 61 PVGLVVCVPLGASDHWGRYGRRRPFIWALSGLILLSLFIIPRAGWLAGLLCPDPRPL 120

QY 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROQVSVYAFMISLGGCLGYLLPA 180  
DB 121 ELALLILGVGLDFCGQVCFPLEALLSDFRDPDHCROQVSVYAFMISLGGCLGYLLPA 180

QY 181 IDWDTSALAPYLGTQBECLFGLITLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240  
DB 181 IDWDTSALAPYLGTQBECLFGLITLFTCTVAATLLVAEEAALGPTPEAGLSAPLSPH 240

QY 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300  
DB 241 CCPCARLAFRNLGALLPRHLQCCMRPTRLRLFVAELCSNMALMTFTLFTYDFVGEGL 300

QY 301 YQGVPAEPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360  
DB 301 YQGVPAEPGTEARRHYDEGVMSGLFLQCAISLVFSLVMDRLVQRTGTRAVYLASVA 360

QY 361 APPVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTG 420  
DB 361 APPVAAGATCLSHSVAVVTASALTGTFTSALQILPYTLASLYHREKQVFLPKYRGDTG 420

QY 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGGLPPPPALCGASACDVSVRVVVGEPTEA 480  
DB 421 ASSEDSLMTSFLPGPKGAPFPNGHVAGSGGLPPPPALCGASACDVSVRVVVGEPTEA 480

QY 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAYIFATQ 540  
DB 481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMVSAAGLGLVAYIFATQ 540

QY 541 VVFDKSDLAKYSA 553  
DB 541 VVFDKSDLAKYSA 553

Search completed: December 3, 2003, 17:25:48  
Job time : 37 secs

RESULT 15

US-10-010-940-113  
; Sequence 113, Application US/10010940  
; Publication No. US20030088062A1  
; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun  
; APPLICANT: Dillon, Davin C.  
; APPLICANT: Mitcham, Jennifer L.  
; APPLICANT: Harlocker, Susan Louise  
; APPLICANT: Jiang Yuqi  
; APPLICANT: Reed, Steven G.  
; APPLICANT: Kalos, Michael  
; APPLICANT: Fanger, Gary  
; APPLICANT: Retter, Mark  
; APPLICANT: Solk, John  
; APPLICANT: Day, Craig

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND

; FILE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER

; FILE REFERENCE: 210121.427D3

; CURRENT APPLICATION NUMBER: US/10/010,940

; CURRENT FILING DATE: 2001-12-05

; NUMBER OF SEQ ID NOS: 575

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 113

; LENGTH: 553

; TYPE: PRT

; ORGANISM: Homo sapien

US-10-010-940-113

Query Match 100.0%; Score 553; DB 15; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVORLWVSRLRRKQAQLLLVNLITGLEVCLAAAGITYPPLLLEVGVEEKFMTWVLGIG 60
Db	1	MVORLWVSRLRRKQAQLLLVNLITGLEVCLAAAGITYPPLLLEVGVEEKFMTWVLGIG 60
Qy	61	PVLGLVCVPLGSGASDHWGRYGRRRPFIWALSIGLILLSLFLIPRAGWLAGLLCPDPRPL 120
Db	61	PVLGLVCVPLGSGASDHWGRYGRRRPFIWALSIGLILLSLFLIPRAGWLAGLLCPDPRPL 120
Qy	121	ELALLILGVGLDFCGQVCFPTLEALLSDFRDPDHCROAYSVYAFWISLGGCLGYLLPA 180
Db	121	ELALLILGVGLDFCGQVCFPTLEALLSDFRDPDHCROAYSVYAFWISLGGCLGYLLPA 180
Qy	181	IDWDTGALAPYLGTQBECLFGLITLIFLTCVAATLLVABEAAALGPTPEAGLSAPLSPH 240
Db	181	IDWDTGALAPYLGTQBECLFGLITLIFLTCVAATLLVABEAAALGPTPEAGLSAPLSPH 240
Qy	241	CCPCRLAFRNIGALLPRHLQCCRMPTRELFLVAELCSWMLMTFTLFYTDFFVGEGL 300
Db	241	CCPCRLAFRNIGALLPRHLQCCRMPTRELFLVAELCSWMLMTFTLFYTDFFVGEGL 300
Qy	301	YQGVPRAEPTGEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
Db	301	YQGVPRAEPTGEARRHYDEGRVMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
Qy	361	APFVAAGATCLSHSVAVVTASAAITGFTFSALQILPYTLTSLVHREKQVFLPKYRGDTGG 420
Db	361	APFVAAGATCLSHSVAVVTASAAITGFTFSALQILPYTLTSLVHREKQVFLPKYRGDTGG 420
Qy	421	ASSEDSLMTSEFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
Db	421	ASSEDSLMTSEFLPGPKGAPFPNGHVAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
Qy	481	RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIYFATQ 540
Db	481	RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIQVLSQSVTAYMWSAAGLGLVAIYFATQ 540
Qy	541	VVPDKSDLAKYSA 553
Db	541	VVPDKSDLAKYSA 553

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c), 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:20:47 ; Search time 40 Seconds  
(without alignments)

**Title:** US-09-593-793A-113

Perfect score:

Sequence: 1 MVQRLWVSRLLRHRKAQLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: OLIGO

Scoring table: CHICO Gapop 60.0 , Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size : 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Maximum DB seq length: 2000000000

### Post-processing: Listing first 45 summaries

**Database :**

**SPTREMBL 23:\***

```

1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvius:*
16: sp_bacteriap:*
17: sp_archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	553	100.0	553	4	Q96JT2	Q96jt2 homo sapien
2	158	28.6	553	6	Q95KI5	Q95ki5 macaca fasc
3	131	23.7	501	6	Q95KC5	Q95kc5 macaca fasc
4	59	10.7	450	11	Q8K252	Q8k252 mus musculus
5	59	10.7	553	11	Q8KH7	Q8kh7 mus musculus
6	33	6.0	112	11	Q8RI10	Q8ri10 mus musculus
7	10	1.8	520	4	Q8NFH8	Q8nfh8 homo sapien
8	10	1.8	521	4	Q43428	Q43428 homo sapien
9	10	1.8	659	4	Q8NFI5	Q8nfi5 homo sapien
10	9	1.6	305	16	Q8DL01	Q8dl01 synechococ
11	9	1.6	488	10	Q3FV32	Q3fv32 solanum tub
12	9	1.6	500	10	Q3FV32	Q3fv32 lycopersicon
13	9	1.6	631	16	Q8E924	Q8e924 shewanella
14	8	1.4	68	15	Q9QJC6	Q9qjc6 human immun
15	8	1.4	115	15	Q90UA8	Q90ua8 human immun
16	8	1.4	120	2	Q93PB7	Q93pb7 microscilla

17	8	1.4	135	10	Q9LTS6
18	8	1.4	135	10	Q8LDI6
19	8	1.4	142	16	Q8YVF9
20	8	1.4	176	5	Q94075
21	8	1.4	194	10	Q9LRT3
22	8	1.4	232	16	Q8R8K3
23	8	1.4	234	17	Q8TXI9
24	8	1.4	242	2	Q9XEN0
25	8	1.4	260	2	Q543B6
26	8	1.4	282	16	Q8DBF6
27	8	1.4	315	16	Q98G10
28	8	1.4	354	2	Q59257
29	8	1.4	378	10	Q942N6
30	8	1.4	404	10	Q9C3D2
31	8	1.4	405	16	Q8E320
32	8	1.4	405	16	Q8DYC6
33	8	1.4	412	16	Q9A686
34	8	1.4	445	10	Q9CGA1
35	8	1.4	457	2	Q9F499
36	8	1.4	474	10	Q04516
37	8	1.4	480	16	Q05438
38	8	1.4	498	16	Q9KZ81
39	8	1.4	501	10	Q065929
40	8	1.4	510	10	Q9FE59
41	8	1.4	510	10	Q9W3R4
42	8	1.4	594	10	Q080605
43	8	1.4	596	16	Q9A2N0
44	8	1.4	604	10	Q9FVL6
45	8	1.4	607	10	Q8LPM4

## ALIGNMENTS

## RESULT 1

Q96JT2 ID Q96JT2 PRELIMINARY: PRT: 553 AA.

Query Match	100.0%;	Score 553;	DB 4;	Length 553;
Best Local Similarity	100.0%;	Pred. NO. 0;		
Matches 553;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 MVQRLWYSRLRLRRKQAQLLLVLLITFGLEVCLAAGITYVPPILLLEVGVEEKFTMTVLGIG 60

Db 1 MVQRLWYSRLRLRRKQAQLLLVLLITFGLEVCLAAGITYVPPILLLEVGVEEKFTMTVLGIG 60

**Qy**

61 PVLGLVCVPLGSSADHWGRGYRRRPFTWALSIGLILSLFLIPRAGWLACLLCPDRPL 120  
| | | | | | | | | | | | | | | | | | | | |  
**Db**

61 PVLGLVCVPLGSSADHWGRGYRRRPFTWALSIGLILSLFLIPRAGWLACLLCPDRPL 120  
| | | | | | | | | | | | | | | | | | | | |

**Qy**

121 ELALLILGVLLDFCGQCFTPIEALLSDLFRDPDHCROAYSVYAFMISIGGCLGYLLPA 180  

**Db**

121 ELALLILGVLLDFCGQCFTPIEALLSDLFRDPDHCROAYSVYAFMISIGGCLGYLLPA 180  

```

QY 181 IDWTSALAPYLGTOBCECLFGLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240
DB 181 IDWTSALAPYLGTOBCECLFGLTLFLTCVAATLLVAEEAALGPTPEAGLSAPLSPH 240
QY 241 CCPCRLARLAFRNIGALLPRHLQOLCCMRPTRLRLFVAELCSWMLMTFTLFYDFVGEGL 300
DB 241 CCPCRLARLAFRNIGALLPRHLQOLCCMRPTRLRLFVAELCSWMLMTFTLFYDFVGEGL 300
QY 301 YQGVPAEPGTEARRHYDEGVGMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
DB 301 YQGVPAEPGTEARRHYDEGVGMSGLFLQCAISLVFSLVMDRLVQRFCTRAVYLASVA 360
QY 361 AFVAAGATCLSHVAVVTSAAALGTFTFSAILOILPYTLASLYHREKQVLPKYRGDTGG 420
DB 361 AFVAAGATCLSHVAVVTSAAALGTFTFSAILOILPYTLASLYHREKQVLPKYRGDTGG 420
QY 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPTPEA 480
DB 421 ASSEDSLMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPTPEA 480
QY 481 RVVPGRGICLDLAILDSAPLSSQVAPSLFMSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
DB 481 RVVPGRGICLDLAILDSAPLSSQVAPSLFMSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
QY 541 VVFDKSDLAKYSA 553
DB 541 VVFDKSDLAKYSA 553

RESULT 2
Q95K15
ID Q95K15 PRELIMINARY; PRT; 553 AA.
AC Q95K15;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DE Hypothetical 59.4 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Temporal cortex;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060851; BAB46871.1; -
KW Hypothetical protein.
SQ SEQUENCE 553 AA; 59392 MW; 0718F3A91PB3BF1E CRC64;

Query Match 28.6%; Score 158; DB 6; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.8e-144;
Matches 158; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 VPLGSGASHWRGRRRPFIFWALSIGLLSLFLIPRAGWLAGLCPDPRPLEALLIL 127
DB 68 VPLGSGASHWRGRRRPFIFWALSIGLLSLFLIPRAGWLAGLCPDPRPLEALLIL 127
QY 128 GVGLLDFCGQVCTPTEALLSDLFDPDHCQAYSVAFMISLGGCGLYLLPAIDWDTSA 187
DB 128 GVGLLDFCGQVCTPTEALLSDLFDPDHCQAYSVAFMISLGGCGLYLLPAIDWDTSA 187
QY 188 LAPYLGTQBECLFGLTLFLTCVAATLLVAEEAALGP 225
DB 188 LAPYLGTQBECLFGLTLFLTCVAATLLVAEEAALGP 225

RESULT 3

```

```

Q95KC5
ID Q95KC5 PRELIMINARY; PRT; 501 AA.
AC Q95KC5;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DE Hypothetical 53.4 kDa protein.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Medulla oblongata;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB062977; BAB60745.1; -
KW Hypothetical protein.
SQ SEQUENCE 501 AA; 53447 MW; 8C554BBD04EE0470 CRC64;

Query Match 23.7%; Score 131; DB 6; Length 501;
Best Local Similarity 100.0%; Pred. No. 7.6e-118;
Matches 131; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 422 SSEDLSMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPTPEAR 481
DB 422 SSEDLSMTSFLPGPKPGAPFNGHVAGGSGLLPPPPALCGASACDVSVRVVVGPTPEAR 481
QY 430 VVPGRGICLDLAILDSAPLSSQVAPSLFMSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 489
DB 430 VVPGRGICLDLAILDSAPLSSQVAPSLFMSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 489
QY 542 VVFDKSDLAKYS 552
DB 542 VVFDKSDLAKYS 550

RESULT 4
Q8K252
ID Q8K252 PRELIMINARY; PRT; 450 AA.
AC Q8K252;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE Hypothetical protein (Fragment).
GN 2210413P12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC034084; AAH34084.1; -
DR MGD; MGI:1922082; 2210413P12RIK.
DR InterPro; IPR005805; Rieseke.
DR PROSITE; PS00200; RIESEKE_2; 1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 450 AA; 48261 MW; 72C703AEC131302C CRC64;

Query Match 10.7%; Score 59; DB 11; Length 450;
Best Local Similarity 100.0%; Pred. No. 3.8e-48;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 QLCRCMPTRLRLFVAELCSWMLMTFTLFYDFVGEGLYQGVPAEPGTEARRHYDEG 320
DB 262 QLCRCMPTRLRLFVAELCSWMLMTFTLFYDFVGEGLYQGVPAEPGTEARRHYDEG 320

```



```
RESULT 5
Q8KOH7 ID Q8KOH7 PRELIMINARY; PRT; 553 AA.
AC Q8KOH7
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein (PROTEIN homolog).
GN 2210413P12RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; BC031381; AAH31381.1; -.
DR EMBL; AK035428; BAC29063.1; -.
DR MGD; MGI:1922082; 2210413P12RIK.
DR InterPro; IPR005805; Rieske.
DR PROSITE; PS00200; RIESKE_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 553 AA; 59742 MW; 6A474FEA630B4B13 CRC64;

Query Match 10.7%; Score 59; DB 11; Length 553;
Best Local Similarity 100.0%; Pred. No. 4.6e-48;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 SLLLRHKAQLLVNLLTFGLVCLAAAGITVPPPLLEVGVEKFTWVLGIGPVLGLV 66
Db 8 SLLLRHKAQLLVNLLTFGLVCLAAAGITVPPPLLEVGVEKFTWVLGIGPVLGLV 66

RESULT 6
Q8RI10 ID Q8RI10 PRELIMINARY; PRT; 112 AA.
AC Q8RI10
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 11.4 kDa protein (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024519; AAH24519.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 112 AA; 11393 MW; B9D9DE79FEF09FB7 CRC64;

Query Match 6.0%; Score 33; DB 11; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.7e-23;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 485 GRGICLDLAFLDSQVAPSLFNGSIVQLS 517
```

```
Db 44 GRGICLDLAFLDSQVAPSLFNGSIVQLS 76

RESULT 7
Q8NFH8 ID Q8NFH8 PRELIMINARY; PRT; 520 AA.
AC Q8NFH8
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RALBP1 associated Eps domain containing protein 2b.
GN REPS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Oosterhoff J.K.; Penninkhof F.; Brinkmann A.O.; Grootegeod A.J.;
RA Blok L.J.;
RT "Role of REPS2/POB1 in progression of prostate cancer.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF512951; AAM43953.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15_homology.
DR Pfam; PF00036; ehand; 1.
DR SMART; SM00027; EH; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS50031; EH; 1.
SQ SEQUENCE 520 AA; 57773 MW; C5B4F557D58A444D CRC64;

Query Match 1.8%; Score 10; DB 4; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 SGLPPPPPAL 460
Db 369 SGLPPPPPAL 378

RESULT 8
Q43428 ID Q43428 PRELIMINARY; PRT; 521 AA.
AC Q43428
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RalBP1-interacting protein.
GN POB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9086250; PubMed=9422736;
RA Ikeda M.; Ishida O.; Hinoi T.; Kishida S.; Kikuchi A.;
RT "Identification and characterization of a novel protein interacting
RT with Ral-binding protein 1, a putative effector protein of Ral.";
RL J. Biol. Chem. 273:814-821(1998).
DR EMBL; AF010233; AAC02901.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15_homology.
DR Pfam; PF00036; ehand; 1.
DR SMART; SM00027; EH; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PS50031; EH; 1.
SQ SEQUENCE 521 AA; 57901 MW; F5FCF51B56961026 CRC64;

Query Match 1.8%; Score 10; DB 4; Length 521;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 451 SGLPPPPAL 460
Db 370 SGLPPPPAL 379

RESULT 9
O8NF15 PRELIMINARY; PRT; 559 AA.
ID Q8NF15
AC Q8NF15
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE RALBP1 associated Eps domain containing protein 2a.
GN REPS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Oosterhoff J.K., Penninkhof F., Brinkmann A.O., Grootegeed A.J.,
RA Blok L.J.;
RL "Role of REPS2/POB1 in progression of prostate cancer.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF511533; AA43933.1; -.
DR Genew; HGNC:9963; REPS2.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; EPS15_homology.
DR Pfam; PF00036; efhand; 1.
DR SMART; SM00027; EH; 1.
DR PROSITE; PS00018; EF_HAND; 1.
DR PROSITE; PSS0031; EH; 2.
DR PROSITE; PSS0031; EH; 2.
SQ SEQUENCE 559 AA; 71405 MW; 527A658BDFB4582B CRC64;

Query Match 1.8%; Score 10; DB 4; Length 659;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 451 SGLPPPPAL 460
Db 508 SGLPPPPAL 517

RESULT 10
Q8DL01 PRELIMINARY; PRT; 305 AA.
ID Q8DL01
AC Q8DL01
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Indole-3-glycerol phosphate synthase.
GN TRPC OR TLL0698.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.;
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
RL ENBL; AP005371; BAC08249.1; -.
KW Complete proteome.
SQ SEQUENCE 305 AA; 33953 MW; A6B394C1A3CE718B CRC64;

Query Match 1.6%; Score 9; DB 16; Length 305;
Best Local Similarity 100.0%; Pred. No. 7.3;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 VAAGATCLS 372
Db 107 VAAGATCLS 115

RESULT 11
Q9FV92 PRELIMINARY; PRT; 488 AA.
ID Q9FV92
AC Q9FV92
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Sucrose transporter SUT4 (Fragment).
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20407148; PubMed=10948254;
RA Weise A., Barker L., Kuehn C., Lalonde S., Buschmann H., Frommer W.B.,
RA Ward J.M.;
RT "A new subfamily of sucrose transporters, SUT4, with low affinity/high
capacity localized in enucleate sieve elements of plants.";
RL Plant Cell 12:1345-1356(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR ENBL; AF237780; AAG25923.1; -.
DR InterPro; IPR005828; Sub transporter..
DR InterPro; IPR005989; Suc/H_symport.
DR Pfam; PF00083; sugar_cr; 1.
DR TIGRFAMs; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
FT NON_TER 1
FT NON_TER 488
SQ SEQUENCE 488 AA; 52840 MW; 0800388914ADAF77 CRC64;

Query Match 1.6%; Score 9; DB 10; Length 488;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 320 GVRGSLGL 328
Db 319 GVRGSLGL 327

RESULT 12
Q9FVJ6 PRELIMINARY; PRT; 500 AA.
ID Q9FVJ6
AC Q9FVJ6
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Sucrose transporter.
GN SUT4.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Moneymaker;
RX MEDLINE=20407148; PubMed=10948254;
RA Weise A., Barker L., Kuhn C., Lalonde S., Buschmann H., Frommer W.B.,
RA Ward J.M.;
RT "A new subfamily of sucrose transporters, SUT4, with low affinity/high
capacity localized in enucleate sieve elements of plants.";
RL Plant Cell 12:1345-1355(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR ENBL; AF176950; AAG09270.1; -.

```

```
DR InterPro; IPR005828; Sub_transporter.
DR InterPro; IPR005989; Suc/H_symport.
DR Pfam; PF00083; sugar tr; 1.
DR TIGRPFAMS; TIGR01301; GPH_sucrose; 1.
KW Transmembrane.
SQ SEQUENCE 500 AA; 54274 MW; BDFAE7468DA0110 CRC64;

Query Match 1.6%; Score 9; DB 10; Length 500;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 320 GVRMGSLGL 328
Db 326 GVRMGSLGL 334

RESULT 13
Q8E9Z4 PRELIMINARY; PRT; 631 AA.
AC Q8E9Z4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE MSHA biogenesis protein MshH.
GN MSHH OR SO4116.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadales; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward M., Methe B., Clayton R.A.,
RA Deboy R., Teapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA Madupu R., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Vamathevan J., Weidman J.D., Umayam L.A., White O., Wolf A.M.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., Berry K., Lee C.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neillson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015843; AAN57089.1; -.
DR TIGR; SO4116; -.
KW Complete proteome.
SQ SEQUENCE 631 AA; 72176 MW; 8BE730F9D77E4FB1 CRC64;

Query Match 1.6%; Score 9; DB 16; Length 631;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 124 LLILGVGLL 132
Db 130 LLILGVGLL 138

RESULT 14
Q9QJC6 PRELIMINARY; PRT; 68 AA.
AC Q9QJC6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope protein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=151N;

RX MEDLINE=20170331; PubMed=10708058;
RA Casado C., Urtasun I., Martin-Walther M.V., Garcia S., Rodriguez C.,
RA del Romero J., Lopez-Galindez C.;
RT "Genetic analysis of HIV-1 samples from Spain.";
RL J. Acquir. Immune Defic. Syndr. 23:68-74(2000).
DR EMBL; AF152836; AAF08474.1; -.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 68
FT NON_TER 68 68
SQ SEQUENCE 68 AA; 7538 MW; 76B0F5F1734650C7 CRC64;

Query Match 1.4%; Score 8; DB 15; Length 68;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 513 IVQLSQSV 520
Db 3 IVQLSQSV 10

RESULT 15
Q9QUA8 PRELIMINARY; PRT; 115 AA.
ID Q9QUA8
AC Q9QUA8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Envelope glycoprotein (Fragment).
GN ENV.
OS Human immunodeficiency virus 1.
OC Viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21192328; PubMed=11294662;
RA Gunthard H.F., Havlik D.V., Fiscus S., Zhang Z.-Q., Eron J.,
RA Mellors J., Gulick R., Frost S.D., Leigh Brown A.J., Schleif W.,
RA Valentine F., Jonas L., Meibohm A., Ignacio C.C., Isaacs R.,
RA Gamagami R., Emini E., Haase A., Richman D.D., Wong J.K.;
RT "Residual human immunodeficiency virus (HIV) type 1 RNA and DNA in
RT lymph nodes and HIV RNA in genital secretions and in cerebrospinal
RT fluid after suppression of viremia for 2 years.";
RL J. Infect. Dis. 183:1318-1327(2001).
DR EMBL; AF337292; AAKS6214.1; -.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1 115
FT NON_TER 115 115
SQ SEQUENCE 115 AA; 12907 MW; 4490635198880152 CRC64;

Query Match 1.4%; Score 8; DB 15; Length 115;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 513 IVQLSQSV 520
Db 38 IVQLSQSV 45

Search completed: December 3, 2003, 17:24:25
Job time : 41 secs
```

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 3, 2003, 17:21:22 ; Search time 21 seconds  
(without alignments)  
2532.442 Million cell updates/sec

Title: US-09-593-793A-113

Perfect score: 553

Sequence: 1 MVQRLVMSRLRRKQALL.....AIYFATQVVFVDSKLAKYSA 553

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 9616862 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 76.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.6	1609	2 E87243	probable cation tr
2	9	1.6	1632	2 C70752	probable ctpi prot
3	8	1.4	137	2 S16222	bombinin H precurs
4	8	1.4	142	2 AE2058	hypothetical prote
5	8	1.4	176	2 T25087	hypothetical prote
6	8	1.4	223	2 JX0222	ubiquitin thiolest
7	8	1.4	260	1 JS0635	rRNA (adenine-N6-)
8	8	1.4	328	2 A48971	1S1380 polypeptide
9	8	1.4	354	1 S51779	endo-1,4-beta-xyla
10	8	1.4	412	2 G87522	hypothetical prote
11	8	1.4	445	2 C96769	unknown protein F9
12	8	1.4	455	1 WQBST	phosphotransferase
13	8	1.4	456	2 S62331	phosphotransferase
14	8	1.4	474	2 A86234	hypothetical prote
15	8	1.4	480	2 D70600	probable polynucle
16	8	1.4	501	2 T14339	sucrose-proton tra
17	8	1.4	594	2 G84441	probable sucrose-p
18	8	1.4	596	2 E87686	succinate dehydrog
19	8	1.4	787	2 S47245	ActA protein - Lis
20	8	1.4	1080	2 T43164	lactA protein - Li
21	8	1.4	1804	2 T34518	nestin - golden ha
22	7	1.3	68	2 AD3215	hypothetical prote
23	7	1.3	83	2 A71471	hypothetical prote
24	7	1.3	87	2 G46449	hypothetical prote
25	7	1.3	89	1 CCS66	cytochrome c6 - Sp
26	7	1.3	102	2 E87453	hypothetical prote
27	7	1.3	122	2 T49358	related to glycine
28	7	1.3	128	2 G72807	gp64 protein - Myc
29	7	1.3	130	2 S74529	hypothetical prote

thymic shared anti  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
conserved hypothet  
conserved hypothet  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
pupi protein - Pse  
probable arok prot

## ALIGNMENTS

## RESULT 1

E87243

probable cation transport ATPase ML2671 [imported] - Mycobacterium leprae  
C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: E87243

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor

R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,

eam, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq

A;Title: Massive gene decay in the leprosy bacillus

A;Reference number: A86909; MUID:21128732; PMID:11234002

A;Accession: E87243

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1609 &lt;STO&gt;

A;Cross-references: GB:AL450380; NID:G13093871; PIDN:CAC32203.1; GSPDB:GN00147

C;Genetics:

A;Gene: ML2671

Query Match 1.6%; Score 9; DB 2; Length 1609;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 QLLLVNLLT 25

Db 1425 QLLLVNLLT 1433

## RESULT 2

C70752

probable ctpi protein - Mycobacterium tuberculosis (strain H37RV)

C;Species: Mycobacterium tuberculosis

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C;Accession: C70752

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: C70752

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1632 &lt;COL&gt;

A;Cross-references: GB:Z74410; GB:AL123456; NID:G3261600; PIDN:CAA98940.1; PID:e1299890;

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: ctpi

C;Superfamily: ATPase nucleotide-binding domain homology

F;1237-1406/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 1.6%; Score 9; DB 2; Length 1632;  
Best Local Similarity 100.0%; Pred. No. 8.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 QLLVNLTLT 25

|||||

Db 1430 QLLVNLTLT 1438

# RESULT 3

S16222

bombinin H precursor - yellow-bellied toad

N;Contains: Bombinin H

C;Species: Bombina variegata (yellow-bellied toad)

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 21-Jul-2000

C;Accession: S16222; A44581

R;Simmaco, M.; Barra, D.; Chiarini, F.; Noviello, L.; Melchiorri, P.; Kreil, G.; Richters

Eur. J. Biochem. 199, 217-222, 1991

A;Title: A family of bombinin-related peptides from the skin of Bombina variegata.

A;Reference number: S16222; MUID:91293126; PMID:1712299

A;Accession: S16222

A;Molecule type: mRNA

A;Residues: 1-137 <SIM>

A;Cross-references: GB:X59695; NID:962568; PIDN:CAA42216.1; PID:962569

R;Mignogna, G.; Simmaco, M.; Kreil, G.; Barra, D.

EMBO J. 12, 4829-4832, 1993

A;Title: Antibacterial and haemolytic peptides containing D-alloisoleucine from the skin

A;Reference number: S39612; MUID:94038967; PMID:8223491

A;Accession: A44581

A;Status: preliminary

A;Molecule type: protein

A;Residues: 117-136 <MIG>

C;Superfamily: bombinin H precursor

C;Keywords: amidated carboxyl end; D-amino acid

F;117-136/Product: bombinin H #status experimental <MAT>

F;118/Modified site: D-allo-isoleucine (Ile) (in mature form) #status experimental

F;136/Modified site: amidated carboxyl end (Ile) (amide in mature form from following gl

Query Match 1.4%; Score 8; DB 2; Length 137;

Best Local Similarity 100.0%; Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 59 IGPVLGLV 66

|||||

Db 118 IGPVLGLV 125

# RESULT 4

AE2058

hypothetical protein alr2019 [imported] - Nostoc sp. (strain PCC 7120)

C;Species: Nostoc sp. PCC 7120

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Dec-2002

C;Accession: AE2058

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AE2058

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-142 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA873718.1; PID:gl7131109; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: alr2019

Query Match

Best Local Similarity 1.4%; Score 8; DB 2; Length 142;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 LLSLFLIP 104

|||||

Db 86 LLSLFLIP 93

# RESULT 5

T25087

hypothetical protein T21H8.5 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 04-Mar-2000

C;Accession: T25087

R;Gajadaty, S.

submitted to the EMBL Data Library, August 1996

A;Reference number: Z19979

A;Accession: T25087

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-176 <WIL>

A;Cross-references: EMBL:Z78546; PIDN:CAB01770.1; GSPDB:GN00028; CESP:T21H8.5

A;Experimental source: clone T21H8

C;Genetics:

A;Gene: CESP:T21H8.5

A;Map position: X

A;Introns: 25/1; 83/1; 112/1

C;Superfamily: Caenorhabditis elegans hypothetical protein T21H8.5

Query Match

Best Local Similarity 1.4%; Score 8; DB 2; Length 176;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 453 LLPPPPAL 460

|||||

Db 51 LLPPPPAL 58

# RESULT 6

JX0222

ubiquitin thioesterase (EC 3.1.2.15) PCP9.5 - rat

N;Alternate names: ubiquitin carboxy-terminal

C;Species: Rattus norvegicus (Norway rat)

C;Date: 09-Oct-1992 #sequence\_revision 09-Oct-1992 #text\_change 10-Dec-1999

C;Accession: JX0222

R;Kajimoto, Y.; Hashimoto, T.; Shirai, Y.; Nishino, N.; Kuno, T.; Tanaka, C.

J. Biochem. 112, 28-32, 1992

A;Title: cDNA cloning and tissue distribution of a rat ubiquitin carboxyl-terminal hydroly

A;Reference number: JX0222; MUID:93054416; PMID:1331034

A;Accession: JX0222

A;Molecule type: mRNA

A;Residues: 1-223 <KAJ>

A;Cross-references: DDBJ:D01109

A;Experimental source: brain

C;Superfamily: human ubiquitin thioesterase

C;Keywords: thiolester hydrolase

Query Match

Best Local Similarity 1.4%; Score 8; DB 2; Length 223;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 GASSRDSL 427

|||||

Db 186 GASSRDSL 193

# RESULT 7

JS0635

rRNA (adenine-N6-) methyltransferase (EC 2.1.1.48) - Streptomyces lividans

C;Species: Streptomyces lividans

C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 05-May-2000

C;Accession: JS0635

R;Jenkins, G.; Cundliffe, E.

Gene 108, 55-62, 1991

A;Title: Cloning and characterization of two genes from Streptomyces lividans that confer

A;Reference number: JS0635; MUID:92104506; PMID:1761231

A;Accession: JS0635

A;Molecule type: DNA

A;Residues: 1-260 <JUN>

A;Cross-references: GB:W74717; NID:gl53345; PIDN:AAA26779.1; PID:gl53346

A;Experimental source: strain TK21

C;Comment: This enzyme confers resistance to lincomycin.

C;Comment: This enzyme catalyzes the monomethylation of a specific adenosine within 23S

C;Genetics:

A;Gene: lrm

C;Superfamily: rRNA (adenine-N6-)-methyltransferase

C;Keywords: methyltransferase; S-adenosylmethionine

Query Match 1.4%; Score 8; DB 1; Length 260;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 PPLLEVG 47

Db 36 PPLLEVG 43

#### RESULT 8

A48971 IS1380 polypeptide homolog - Bradyrhizobium japonicum

C;Species: Bradyrhizobium japonicum

C;Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 23-Mar-1995

C;Accession: A48971

R;Judd, A.K.; Sadowsky, M.J.

Appl. Environ. Microbiol. 59, 1656-1661, 1993

A;Title: The Bradyrhizobium japonicum serocuster 123 hyperreiterated DNA region, HRS1.

A;Reference number: A48971; MUID:93298005; PMID:8390818

A;Accession: A48971

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-328 <JUD>

A;Experimental source: USDA 424

A;Note: sequence extracted from NCBI backbone (NCBIN:133925, NCBIP:133926)

Query Match 1.4%; Score 8; DB 2; Length 328;

Best Local Similarity 100.0%; Pred. No. 22;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 GLSAPSL 238

Db 275 GLSAPSL 282

#### RESULT 9

S51779 endo-1,4-beta-xylanase (EC 3.2.1.8) Y precursor - Bacillus sp. (strain YA-335)

N;Alternate names: xylanase Y

C;Species: Bacillus sp.

A;Variety: strain YA-335

C;Date: 15-Jul-1995 #sequence\_revision 22-Nov-1996 #text\_change 16-Jun-2000

C;Accession: S51779; S48127

R;Hyun Ju, Y.

submitted to the EMBL Data Library, December 1990

A;Reference number: S51779

A;Accession: S51779

A;Molecule type: DNA

A;Residues: 1-354 <HYU>

A;Cross-references: EMBL:X59059; NID:g487722; PIDN:CAA41784.1; PID:gl334251

A;Experimental source: strain YA-335

R;Yu-Hyun, Y.; Park, Y.S.; Yum, D.Y.; Kim, J.M.; Kong, I.S.; Bai, D.H.

J. Microbiol. Biotechnol. 3, 139-145, 1993

A;Title: Nucleotide sequence and analysis of a xylanase gene (xynS) from alkali-tolerant

A;Reference number: S48126

A;Accession: S48127

A;Molecule type: DNA

A;Residues: 1-17,19-229 <JUH>

A;Cross-references: EMBL:X59059

A;Experimental source: strain YA-335

A;Note: the authors translated the initiation codon TTG for residue 1 as Leu

C;Genetics:

A;Gene: xynY

A;Start codon: TTG

C;Function:

A;Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans

A;Pathway: xylan degradation

C;Superfamily: Clostridium endo-1,4-beta-xylanase B; endo-1,4-beta-xylanase homology

C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation

F;1-27/Domain: signal sequence #status predicted <SIG>

F;28-354/Product: endo-1,4-beta-xylanase Y #status predicted <WAT>

F;39-223/Domain: endo-1,4-beta-xylanase homology <XYL>

F;120,210/Active site: Glu #status predicted

Query Match 1.4%; Score 8; DB 1; Length 354;

Best Local Similarity 100.0%; Pred. No. 23;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 53 MTWVLGIG 60

Db 10 MTWVLGIG 17

#### RESULT 10

G87522

hypothetical protein CC2208 [imported] - Caulobacter crescentus

C;Species: Caulobacter crescentus

C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001

C;Accession: G87522

R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.F.

n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapero, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A;Title: Complete Genome Sequence of Caulobacter crescentus.

A;Reference number: A87249; MUID:21173698; PMID:11259647

A;Accession: G87522

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-412 <STO>

A;Cross-references: GB:AE005673; NID:gl3423711; PIDN:AAK24179.1; GSPDB:GN00148

C;Genetics:

A;Gene: CC2208

Query Match 1.4%; Score 8; DB 2; Length 412;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 LFGLLTLLI 206

Db 80 LFGLLTLLI 87

#### RESULT 11

C96769

unknown protein F9E11.3 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C;Accession: C96769

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: C96769

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-445 <STO>

A;Cross-references: GB:AE005173; NID:g10092422; PIDN:AAG12827.1; GSPDB:GN00141  
 C;Genetics: 1  
 A;Gene: F9ell1.3  
 A;Map position: 1

Query Match 1.4%; Score 8; DB 2; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 93 SLGILLSL 100  
 |||||  
 Db 79 SLGILLSL 86

## RESULT 12

## WOBEST

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific, factor II - Salmonella typhimurium  
 N;Alternate names: phosphotransferase system enzyme II-sucrose; protein- $\pi$ i-phosphohistidine kinase  
 C;Species: Salmonella typhimurium  
 C;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 11-Jun-1999  
 C;Accession: S01036; S62329; S35016  
 R;Ebner, R.; Lengeler, J.W.  
 Mol. Microbiol. 2, 9-17, 1988

A;Title: DNA sequence of the gene scrA encoding the sucrose transport protein enzyme-II (scrA) from *Salmonella typhimurium*.  
 A;Reference number: S01036; MUID:88216186; PMID:3285123  
 A;Accession: S01036  
 A;Molecule type: DNA  
 A;Residues: 1-455.<EBN>  
 A;Cross-references: EMBL:Y00541; NID:g47925; PIDN:CAA68605.1; PID:g47926  
 R;Titgemeyer, F.; Jahreis, K.; Ebner, R.; Lengeler, J.W.  
 Mol. Gen. Genet. 250, 197-206, 1996

A;Title: Molecular analysis of the scrA and scrB genes from *Klebsiella pneumoniae* and *pl* from *Escherichia coli*.  
 A;Reference number: S62329; MUID:96188840; PMID:8628219

A;Accession: S62329  
 A;Molecule type: DNA  
 A;Residues: 1-7,9-24,'HC',26-36,38-46,49-76,78-140,'N','A',145-207,'E',211-230,'V','EKL'  
 R;Jahreis, K.; Lengeler, J.W.  
 Mol. Microbiol. 9, 195-209, 1993

A;Title: Molecular analysis of two ScrR repressors and of a ScrR-FruR hybrid repressor from *Escherichia coli*.  
 A;Reference number: S35014; MUID:94018607; PMID:8412665

A;Accession: S35016  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-24,'HC',26-435,'IAFGVAFTV',445-455<JAH>  
 A;Cross-references: EMBL:X67750; NID:g395262; PIDN:CAA47973.1; PID:g395263  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992

C;Genetics: 1  
 A;Gene: scrA  
 A;Genome: plasmid  
 C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II  
 C;Keywords: membrane protein; phosphoprotein; phosphotransferase; sugar transport system  
 F:308/Active site: His #status predicted

Query Match 1.4%; Score 8; DB 1; Length 455;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 AGWLAGLL 113  
 |||||  
 Db 287 AGWLAGLL 294

## RESULT 13

## S62331

phosphotransferase system enzyme II (EC 2.7.1.69), sucrose-specific - *Klebsiella pneumoniae*  
 C;Species: *Klebsiella pneumoniae*  
 C;Date: 08-Dec-1996 #sequence\_revision 06-Dec-1996 #text\_change 18-Jun-1999  
 C;Accession: S62331; S15195  
 R;Titgemeyer, F.; Jahreis, K.; Ebner, R.; Lengeler, J.W.  
 Mol. Gen. Genet. 250, 197-206, 1996

A;Title: Molecular analysis of the scrA and scrB genes from *Klebsiella pneumoniae* and *pl*

sucrose-6-phosphate invertase.

A;Reference number: S62329; MUID:96188840; PMID:8628219

A;Accession: S62331

A;Molecule type: DNA

A;Residues: 1-456<TIT>

A;Cross-references: EMBL:X57401; NID:g43931; PIDN:CAA40658.1; PID:g43933  
 R;Schmid, K.; Ebner, R.; Jahreis, K.; Lengeler, J.W.; Titgemeyer, F.

Mol. Microbiol. 5, 941-950, 1991

A;Title: A sugar-specific porin, ScrY, is involved in sucrose uptake in enteric bacteria

A;Reference number: S15193; MUID:91312133; PMID:1649946

A;Accession: S15193

A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-456<SCH>

A;Cross-references: EMBL:X57401; NID:g43931; PIDN:CAA40658.1; PID:g43933

A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1991

C;Genetics: 1

A;Gene: scrA

C;Superfamily: phosphotransferase system sucrose-specific enzyme II, factor II

C;Keywords: membrane protein; phosphotransferase; sugar transport system

Query Match 1.4%; Score 8; DB 2; Length 456;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 AGWLAGLL 113  
 |||||  
 Db 288 AGWLAGLL 295

## RESULT 14

## A86234

hypothetical protein [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

A;Accession: A86234

R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Cohen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: A86234

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-474<STO>

A;Cross-references: GB:AE005172; NID:g2160188; PIDN:AAB60751.1; GSPDB:GN00141

C;Genetics: 1

A;Map position: 1

C;Superfamily: common tobacco sucrose transport protein

Query Match 1.4%; Score 8; DB 2; Length 474;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 YGRRRPFI 89  
 |||||  
 Db 107 YGRRRPFI 114

## RESULT 15

## D70600

probable polynucleotide polymerase - *Mycobacterium tuberculosis* (strain H37Rv)

C;Species: *Mycobacterium tuberculosis*

C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C;Accession: D70600

R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;



Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230  
 A;Accession: D70600  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-480 <COL>  
 A;Cross-references: GB:294121; GB:AL123456; NID:g3261736; PIDN:CAB08094.1; PID:e312270;  
 C;Experimental source: strain H37Rv  
 C;Genetics:  
 A;Gene: pcnA

Query Match 1.4%; Score 8; DB 2; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 254 GALLPRLH 261  
 |||||  
 Db 375 GALLPRLH 382

Search completed: December 3, 2003, 17:24:59  
 Job time : 22 secs

**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 3, 2003, 17:19:51 ; Search time 18 seconds  
(without alignments)  
1444.765 Million cell updates/sec

Title: US-09-593-793A-113

Perfect score: 553

Sequence: 1 MVQRLMVSRLRHRKAQLL.....AIYFATQVVPKSLAKYSA 553

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.6	1609	1	CTPI MYCLE
2	9	1.6	1625	1	CTPI MYCTU
3	8	1.4	21	1	BOH4 BOMVA
4	8	1.4	21	1	BOH5 BOMVA
5	8	1.4	137	1	BLP1 BOMVA
6	8	1.4	137	1	BLP2 BOMVA
7	8	1.4	223	1	UBL1 MOUSE
8	8	1.4	223	1	UBL1 RAT
9	8	1.4	456	1	PTSB KLEPN
10	8	1.4	456	1	PTSB SALTY
11	8	1.4	745	1	NLDL RAT
12	8	1.4	1158	1	ALAI ARATH
13	7	1.3	20	1	MAX7 BOMMX
14	7	1.3	78	1	EX7S PARZE
15	7	1.3	89	1	CYC6 SPIMA
16	7	1.3	128	1	VG64 BPM2
17	7	1.3	130	1	LY6E MOUSE
18	7	1.3	144	1	MAX4 BOMMX
19	7	1.3	145	1	SODM STRAI
20	7	1.3	149	1	YQCA ECOLI
21	7	1.3	151	1	YPH7 CHRVI
22	7	1.3	151	1	YQCA ERWCA
23	7	1.3	173	1	YL18 AQUAE
24	7	1.3	176	1	AROK MYCTU
25	7	1.3	182	1	RBS1 MESCR
26	7	1.3	183	1	YM39 MARPO
27	7	1.3	195	1	YBJE HAEIN
28	7	1.3	197	1	PTCA MOUSE
29	7	1.3	200	1	WINI SOLTU
30	7	1.3	206	1	PTCA HUMAN
31	7	1.3	220	1	TEGP HSVE4
32	7	1.3	226	1	COLI ONCKE
33	7	1.3	227	1	YB61 DROME

## RESULT 1

ID	CTPI MYCLE	STANDARD;	PRT;	1609 AA.
AC	OS3114;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Probable cation-transporting ATPase I (EC 3.6.3.-)			
GN	CTPI OR ML2671 OR MLCB1913.02.			
OS	Mycobacterium leprae.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1769;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TN;			
RX	MEDLINE=21128732; PubMed=11234002;			
RA	Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,			
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,			
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,			
RA	Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,			
RA	Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,			
RA	Squires S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,			
RA	Barrell B.G.;			
RT	"Massive gene decay in the leprosy bacillus.";			
RL	Nature 409:1007-1011(2001)			
CC	- - CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.			
CC	- - SUBCELLULAR LOCATION: Integral membrane protein.			
CC	- - SIMILARITY: Belongs to the cation transport ATPases family (P-type ATPases).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	EMBL; AL022118; CAA17934.1; -			
CC	EMBL; AL583926; CAC32203.1; -			
DR	PIR; E87243; E87243.			
DR	Leptoma; ML2671; -			
DR	InterPro; IPR001757; ATPase_E1-E2.			
DR	InterPro; IPR000695; H ATPase.			
DR	InterPro; IPR005834; Hydrolase.			
DR	Pfam; PF00122; E1-E2 ATPase; 1.			
DR	Pfam; PF00702; Hydrolase; 1.			
DR	PRINTS; PR00119; CATAPASE.			
DR	PRINTS; PR00120; HAIPASE.			
DR	TIGRFAMs; TIGR01494; ATPase_P-type; 5.			
DR	PROSITE; PS00154; ATPase E1 E2; 1.			
KW	Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding; Complete proteome.			

Q04618 oncorhynchu  
P42250 bacillus eu  
P46692 gallus gall  
Q95336 homo sapien  
Q9wy77 thermotoga  
Q99848 homo sapien  
Q9d903 mus musculu  
P32740 caenorhabdi  
Q92554 equalus aca  
P37763 neisseria g  
Q03723 saccharomyc  
P41654 methanobact

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

CC EMBL; 274410; CAA98940.1; --  
CC EMBL; AE006923; AAK44338.1; --  
CC TIGR; C70752; C70752.  
CC TIGR; MT0116; --  
CC InterPro; IPR001757; ATPase\_E1-E2.  
CC InterPro; IPR000695; H-ATPase.  
CC InterPro; IPR005834; Hydrolase.  
CC Pfam; PF00122; E1-E2\_ATPase; 1.  
CC Pfam; PF00702; Hydrolase; 1.  
CC PRINTS; PR00119; CATATPASE.  
CC PRINTS; PR00120; HATPASE.  
CC TIGRPFAM; TIGR01494; ATPase\_P-type; 6.  
CC PROSITE; PS00154; ATPase\_E1-E2; 1.  
CC Hydrolase; Transmembrane; Phosphorylation; Magnesium; ATP-binding;  
KW Complete proteome.  
FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 148 168 POTENTIAL.  
FT TRANSMEM 177 197 POTENTIAL.  
FT TRANSMEM 315 335 POTENTIAL.  
FT TRANSMEM 358 378 POTENTIAL.  
FT TRANSMEM 637 657 POTENTIAL.  
FT TRANSMEM 673 693 POTENTIAL.  
FT TRANSMEM 778 798 POTENTIAL.  
FT TRANSMEM 921 941 POTENTIAL.  
FT TRANSMEM 969 989 POTENTIAL.  
FT TRANSMEM 997 1017 POTENTIAL.  
FT TRANSMEM 1401 1421 POTENTIAL.  
FT TRANSMEM 1432 1452 POTENTIAL.  
FT TRANSMEM 1547 1567 POTENTIAL.  
FT MOD\_RES 1053 1053 PHOSPHORYLATION (BY SIMILARITY).  
FT METAL 1340 1340 MAGNESIUM (BY SIMILARITY).  
FT METAL 1344 1344 MAGNESIUM (BY SIMILARITY).  
FT CONFLICT 1571 1625 DERPDADSDAGG -> SGVPRWDRSPGRASSAPRQPPQ  
FT SORWHRSGWQAOVSCLNLMNALTTRKTLTRVDRTVRRPR  
FT (IN REF. 1)  
SQ SEQUENCE 1625 AA; 167805 MW; 54082AD7064C22EB CRC64;

Query Match 1.6%; Score 9; DB 1; Length 1625;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 QLLLVNLLT 25  
Db 1430 QLLLVNLLT 1438  
|||||  
|||||

RESULT 3  
BOH4\_BOMVA  
ID BOH4\_BOMVA STANDARD; PRT; 21 AA.  
AC P82284;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Bombinin H4.  
OS Bombina variegata (Yellow-bellied toad).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.  
OX NCBI\_TaxID=8348;  
RN [1]  
RP SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
RC TISSUE=Skin secretion;  
RX MEDLINE=94038967; PubMed=8223491;  
RA Mignogna G., Simmaco M., Kreil G., Barria D.;  
RT "Antibacterial and haemolytic peptides containing D-alloisoleucine

FT TRANSMEM 30 50 POTENTIAL.  
FT TRANSMEM 176 196 POTENTIAL.  
FT TRANSMEM 238 258 POTENTIAL.  
FT TRANSMEM 357 377 POTENTIAL.  
FT TRANSMEM 641 661 POTENTIAL.  
FT TRANSMEM 673 693 POTENTIAL.  
FT TRANSMEM 778 798 POTENTIAL.  
FT TRANSMEM 921 941 POTENTIAL.  
FT TRANSMEM 969 989 POTENTIAL.  
FT TRANSMEM 997 1017 POTENTIAL.  
FT TRANSMEM 1396 1416 POTENTIAL.  
FT TRANSMEM 1426 1446 POTENTIAL.  
FT TRANSMEM 1542 1562 POTENTIAL.  
FT TRANSMEM 1573 1593 POTENTIAL.  
FT MOD\_RES 1053 1053 PHOSPHORYLATION (BY SIMILARITY).  
FT METAL 1335 1335 MAGNESIUM (BY SIMILARITY).  
FT METAL 1339 1339 MAGNESIUM (BY SIMILARITY).  
SQ SEQUENCE 1609 AA; 166871 MW; 58FA2079905E3395 CRC64;

Query Match 1.6%; Score 9; DB 1; Length 1609;  
Best Local Similarity 100.0%; Pred. No. 3.7;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 QLLLVNLLT 25  
Db 1425 QLLLVNLLT 1433  
|||||  
|||||

RESULT 2  
CTPI\_MYCTU  
ID CTPI\_MYCTU STANDARD; PRT; 1625 AA.  
AC Q1090;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Probable cation-transporting ATPase I (EC 3.6.3.-).  
GN CTPI OR RV0107C OR MT0116 OR MTCV251.26C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RC STRAIN=98295987; PubMed=9634230;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence."  
RL Nature 393:537-544 (1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains."  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type  
ATPases).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

RT from the skin of Bombina variegata.";  
 RL EMO J. 12:4829-4832(1993).  
 CC -1- FUNCTION: HAS ANTIMICROBIAL AND HEMOLYTIC ACTIVITIES.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Skin.  
 CC -1- SIMILARITY: BELONGS TO THE BOMBININ FAMILY.  
 KW Amphibian defense peptide; Antibiotic; Hemolysis; Amidation;  
 KW D-amino acid.  
 FT PEPTIDE 1 20 BOMBININ H4.  
 FT MOD RES 2 2 D-ALLO-ISOLEUCINE.  
 FT MOD RES 20 20 AMIDATION (G-21 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 21 AA; 1975 MW; 08C7281E1D6D43BD CRC64;  
 Query Match 1.4%; Score 8; DB 1; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.98;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 IGPVLGLV 66  
 Db 2 IGPVLGLV 9  
 RESULT 4  
 BOH5\_BOMVA STANDARD; PRT; 21 AA.  
 AC P82285;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bombinin H5.  
 OS Bombina variegata (Yellow-bellied toad).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.  
 OX NCBI\_TaxID=8348;  
 RN [1]  
 RP SEQUENCE, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RC TISSUE=Skin secretion;  
 RA Mignogna G., Simmaco M., Kreil G., Barra D.;  
 RT "Antibacterial and haemolytic peptides containing D-alloisoleucine  
 from the skin of Bombina variegata.";  
 RL EMO J. 12:4829-4832(1993).  
 CC -1- FUNCTION: HAS ANTIMICROBIAL AND HEMOLYTIC ACTIVITIES.  
 CC -1- SIMILARITY: BELONGS TO THE BOMBININ FAMILY.  
 KW Amphibian defense peptide; Antibiotic; Hemolysis; Amidation;  
 KW D-amino acid.  
 FT PEPTIDE 1 20 BOMBININ H5.  
 FT MOD RES 2 2 D-ALLO-ISOLEUCINE.  
 FT MOD RES 20 20 AMIDATION (G-21 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 21 AA; 1975 MW; 08C7281E1BDF3BD CRC64;  
 Query Match 1.4%; Score 8; DB 1; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 0.98;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 IGPVLGLV 66  
 Db 2 IGPVLGLV 9  
 RESULT 5  
 BLP1\_BOMVA STANDARD; PRT; 137 AA.  
 AC P29006;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bombinin-like peptide 1 precursor [Contains: Acidic peptide 1-1;  
 DE Bombinin-like peptide 1 (BLP-1); Octapeptide 1; Acidic peptide 1-2;  
 DE Bombinin H].  
 OS Bombina variegata (Yellow-bellied toad).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.  
 OX NCBI\_TaxID=8348;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PRELIMINARY SEQUENCE OF 44-69.  
 RC TISSUE=Skin, and Skin secretion;  
 RX MEDLINE=91293126; PubMed=1712299;  
 RA Simmaco M., Barra D., Chiarini F., Novello L., Melchiorri P.,  
 RA Kreil G., Richter K.;  
 RT "A family of bombinin-related peptides from the skin of Bombina  
 variegata.";  
 RL Eur. J. Biochem. 199:217-222(1991).  
 RN [2]  
 RP SEQUENCE OF 117-136, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RC TISSUE=Skin secretion;  
 RX MEDLINE=94038967; PubMed=8223491;  
 RA Mignogna G., Simmaco M., Kreil G., Barra D.;  
 RT "Antibacterial and haemolytic peptides containing D-alloisoleucine  
 from the skin of Bombina variegata.";  
 RL EMO J. 12:4829-4832(1993).  
 CC -1- FUNCTION: HAS ANTIMICROBIAL ACTIVITY, BUT NO HEMOLYTIC ACTIVITY.  
 CC PRELIMINARY EVIDENCE INDICATES THAT THIS PEPTIDE DOES NOT LYSO AND  
 THUS KILL THE BACTERIA BY ITS ANTIMICROBIAL ACTIVITY.  
 CC -1- FUNCTION: BOMBININ H HAS ANTIBACTERIAL AND HEMOLYTIC ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Skin.  
 CC -1- SIMILARITY: BELONGS TO THE BOMBININ FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X59695; CAA42216.1;  
 DR PIR; S16222; S16222.  
 DR Pfam; PF05298; Bombinin; 1.  
 KW Amphibian defense peptide; Antibiotic; Amidation;  
 KW Cleavage on pair of basic residues; Signal; Hemolysis; D-amino acid.  
 FT SIGNAL 1 18 POTENTIAL.  
 FT PEPTIDE 19 43 ACIDIC PEPTIDE 1-1 (POTENTIAL).  
 FT PEPTIDE 44 70 BOMBININ-LIKE PEPTIDE 1.  
 FT PEPTIDE 74 81 OCTAPEPTIDE (POTENTIAL).  
 FT PEPTIDE 84 114 ACIDIC PEPTIDE 1-2 (POTENTIAL).  
 FT PEPTIDE 117 136 BOMBININ H.  
 FT MOD RES 70 70 AMIDATION (G-71 PROVIDE AMIDE GROUP).  
 FT MOD RES 118 118 D-ALLO-ISOLEUCINE.  
 FT MOD RES 136 136 AMIDATION (G-137 PROVIDE AMIDE GROUP).  
 FT VARIANT 117 117 I -> L.  
 FT VARIANT 124 124 L -> M.  
 SQ SEQUENCE 137 AA; 14982 MW; 3EC3EF6E47CA1F92 CRC64;  
 Query Match 1.4%; Score 8; DB 1; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 IGPVLGLV 66  
 Db 118 IGPVLGLV 125  
 RESULT 6  
 BLP2\_BOMVA STANDARD; PRT; 137 AA.  
 AC P82286;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bombinin-like peptide 2 precursor [Contains: Acidic peptide 2-1;  
 DE Bombinin-like peptide 2 (BLP-2); Octapeptide 2; Acidic peptide 2-2;  
 DE Bombinin H2].

OS Bombina variegata (Yellow-bellied toad).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.  
 OX NCBI\_TaxID=9348;  
 [1]  
 RN SEQUENCE FROM N.A., AND SEQUENCE OF 117-136.  
 RP TISSUE=Skin, and Skin secretion;  
 RC MEDLINE=99266278; PubMed=10333736;  
 RX Mignogna G., Sannacoe M., Kreil G., Barra D.;  
 RA "Antibacterial and haemolytic peptides containing D-alloisoleucine  
 RT from the skin of Bombina variegata.";  
 RL EMBO J. 12:4829-4832(1993).  
 [2]  
 RN Biopolymers 47:435-450(1998).  
 RP SEQUENCE OF 117-136, AND POST-TRANSLATIONAL MODIFICATIONS.  
 RC TISSUE=Skin secretion;  
 RX MEDLINE=94038967; PubMed=8223491;  
 RA Mignogna G., Sannacoe M., Kreil G., Barra D.;  
 RT "Antibacterial and haemolytic peptides containing D-alloisoleucine  
 from the skin of Bombina variegata.";  
 RL EMBO J. 12:4829-4832(1993).  
 CC -1- FUNCTION: BOMBININ-LIKE PEPTIDE 2 HAS ANTIMICROBIAL ACTIVITY. BUT  
 NO HEMOLYTIC ACTIVITY. PRELIMINARY EVIDENCE INDICATES THAT THIS  
 PEPTIDE DOES NOT LYSIS AND THUS KILL THE BACTERIA BY ITS  
 ANTIMICROBIAL ACTIVITY.  
 CC -1- FUNCTION: BOMBININ H2 HAS ANTIBACTERIAL AND HEMOLYTIC ACTIVITY.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- TISSUE SPECIFICITY: Skin.  
 CC -1- SIMILARITY: BELONGS TO THE BOMBININ FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AJ251565; CAB61443.1; --  
 DR Pfam; PF05298; Bombinin; 1.  
 KW Amphibian defense peptide; Antibiotic; Amidation;  
 KW Cleavage on pair of basic residues; Signal; Hemolysis.  
 FT SIGNAL 1 18  
 FT PEPTIDE 19 43 ACIDIC PEPTIDE 2-1 (POTENTIAL).  
 FT PEPTIDE 44 70 BOMBININ-LIKE PEPTIDE 2.  
 FT PEPTIDE 74 81 OCTAPEPTIDE (POTENTIAL).  
 FT PEPTIDE 84 114 ACIDIC PEPTIDE 2-2 (POTENTIAL).  
 FT PEPTIDE 117 136 BOMBININ H2.  
 FT MOD RES 70 70 AMIDATION (G-71 PROVIDE AMIDE GROUP).  
 FT MOD RES 136 136 AMIDATION (G-137 PROVIDE AMIDE GROUP).  
 SQ SEQUENCE 137 AA; 15035 MW; 6BFBD1A5F0B6CCD6 CRC64;  
 -----  
 Query Match 1.4%; Score 8; DB 1; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 4.7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 59 IGPVLGLV 66  
 DQ 118 IGPVLGLV 125  
 |||||  
 |||||  
 RESULT 7  
 UBL1\_MOUSE  
 ID UBL1\_MOUSE STANDARD; PRT; 223 AA.  
 AC Q9R0F9; Q9R122;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ubl1 Ubiquitin carboxyl-terminal hydrolase isozyme L1 (EC 3.4.19.12) (UCH-  
 DE L1) (Ubiquitin thiolesterase L1) (Neuron cytoplasmic protein 9.5)  
 DE (RGP 9.5) (RGP9.5).  
 GN UCHL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=99400551; PubMed=10471497;  
 RA Saigoh K., Wang Y.-L., Suh J.G., Yamanishi T., Sakai Y., Kiyosawa H.,  
 RA Harada T., Ichihara N., Wakana S., Kikuchi T., Wada K.;  
 RT "Intragenic deletion in the gene encoding ubiquitin carboxy-terminal  
 RT hydrolase in gad mice.";  
 RL Nat. Genet. 23:47-51(1999).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=Pituitary;  
 RA Marzban G., Grillari J., Reisinger E., Hemetseberger T.,  
 RA Hohenwarter O., Katinger H.;  
 RT "Cloning of the mouse homologue of ubiquitin carboxyl-terminal  
 RT hydrolase PGP9.5.";  
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=C57BL/6J; TISSUE=Hippocampus;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,  
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Fleischmann W., Gaasterland T., Gless C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stabli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Tovo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- FUNCTION: UBIQUITIN-PROTEIN HYDROLASE IS INVOLVED BOTH IN THE  
 PROCESSING OF UBIQUITIN PRECURSORS AND OF UBIQUINATED PROTEINS.  
 CC THIS ENZYME IS A THIOL PROTEASE THAT RECOGNIZE AND HYDROLYZE  
 A PEPTIDE BOND AT THE C-TERMINAL GLYCINE OF UBIQUITIN.  
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =  
 ubiquitin + a thiol.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed in brain and testis.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C12.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; AB025313; BAA84083.1; --  
 DR EMBL; AF172334; RAD51029.1; --  
 DR EMBL; AK013729; BAB28976.1; --  
 DR HSP; F15374; UCH.  
 DR MEROPS; C12.001; --  
 DR MGD; MGI:103149; Uchl1.  
 DR InterPro; IPR001578; UCH 1.  
 DR Pfam; PF01088; Peptidase\_C12; 1.  
 DR PRINTS; PR00707; UBCTHYDLASE.  
 DR ProDom; PD350662; UCH 1; 1.  
 DR PROSITE; PS00140; UCH 1; 1.  
 KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.  
 FT ACT\_SITE 90 90 BY SIMILARITY.

FT ACT\_SITE 161 176 176 176 BY SIMILARITY.  
 FT ACT\_SITE 176 176 176 176 UBIQUITIN BINDING 1 (POTENTIAL).  
 FT DOMAIN 37 54 54 54 UBIQUITIN BINDING 2 (POTENTIAL).  
 FT DOMAIN 170 178 178 178 UBIQUITIN BINDING 2 (POTENTIAL).  
 FT CONFLICT 149 149 E -> K (IN REF. 2).  
 SQ SEQUENCE 223 AA; 24838 MW; F1402BF7B0C077EA CRC64;

Query Match 1.4%; Score 8; DB 1; Length 223;  
 Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 GASEDSL 427  
 DB 186 GASEDSL 193

RESULT 8  
 UBL1\_RAT STANDARD; PRT; 223 AA.  
 AC Q00981;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Ubiquitin carboxyl-terminal hydrolase isozyme L1 (EC 3.4.19.12) (UCH-L1) (Ubiquitin thiolesterase L1) (Neuron cytoplasmic protein 9.5) (GCP 9.5) (GCP9.5).  
 DE (GCP 9.5) (GCP9.5).  
 GN UCHL1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93054416; PubMed=1331034;  
 RA Kajimoto Y., Hashimoto T., Shirai Y., Nishino N., Kuno T., Tanaka C.;  
 RT "CDNA cloning and tissue distribution of a rat ubiquitin carboxyl-terminal hydrolase pGp9.5.";  
 RL J. Biochem. 112:28-32(1992).  
 CC -1- FUNCTION: UBIQUITIN-PROTEIN HYDROLASE IS INVOLVED BOTH IN THE PROCESSING OF UBIQUITIN PRECURSORS AND OF UBIQUINATED PROTEINS. THIS ENZYME IS A THIOL PROTEASE THAT RECOGNIZES AND HYDROLYZES A PEPTIDE BOND AT THE C-TERMINAL GLYCINE OF UBIQUITIN.  
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O = ubiquitin + a thiol.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C12.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; D10699; BAA01541.1;  
 DR HSP; P15374; UCH.  
 DR MEROPS; C12.001; -.  
 DR InterPro; IPR001578; UCH 1.  
 DR Pfam; PF01088; Peptidase C12; 1.  
 DR PRINTS; PR00707; UBCTHYDLASE.  
 DR ProDom; PD350662; UCH 1; 1.  
 DR PROSITE; PS00140; UCH 1; 1.  
 DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.  
 KW ACT\_SITE 90 90 BY SIMILARITY.  
 FT ACT\_SITE 161 161 BY SIMILARITY.  
 FT ACT\_SITE 176 176 BY SIMILARITY.  
 FT DOMAIN 37 54 UBIQUITIN BINDING 1 (POTENTIAL).  
 FT DOMAIN 170 178 UBIQUITIN BINDING 2 (POTENTIAL).  
 SQ SEQUENCE 223 AA; 24782 MW; F1B0A4FB5B4625E CRC64;

Query Match 1.4%; Score 8; DB 1; Length 223;

Best Local Similarity 100.0%; Pred. No. 7;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 420 GASEDSL 427  
 DB 186 GASEDSL 193

RESULT 9  
 PTSE\_KLEPN STANDARD; PRT; 456 AA.  
 AC P27219;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-permease IIBC component) (Phosphotransferase enzyme II, BC component) (EC 2.7.1.69) (BII-SCR).  
 DE (EC 2.7.1.69) (BII-SCR).  
 GN SCRA.  
 OS Klebsiella pneumoniae.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Klebsiella.  
 OC NCBI\_TaxID=573;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=1033-SP14 / KAY2026;  
 RX MEDLINE=96188840; PubMed=8628219;  
 RA Titgemeyer F., Jahreis K., Ebner R., Lengeler J.W.;  
 RT "Molecular analysis of the scrA and scrB genes from Klebsiella pneumoniae and plasmid pUR400, which encode the sucrose transport protein Enzyme II Scr of the phosphotransferase system and a sucrose-6-phosphate invertase.";  
 RT sucrose-6-phosphate invertase.";  
 RL Mol. Gen. Genet. 250:197-206(1996).  
 CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHENOLPYRUVATE-DEPENDENT SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY PHOSPHORYLATION SITE (THE DONOR IS PHOSHO-HPR); IIA TRANSFERS ITS PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A COVALENTLY BOUND EIIB DOMAIN. INSTEAD, EIIB-SCR-MEDIATED PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.  
 CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein histidine + sugar phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 CC -1- SIMILARITY: Contains 1 PTS EIIB domain.  
 CC -1- SIMILARITY: Contains 1 PTS EIIC domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X57401; CAA40658.1; -.  
 DR PIR; S62331; S62331.  
 DR InterPro; IPR001996; PTS EIIB.  
 DR InterPro; IPR003352; PTS EIIC.  
 DR InterPro; IPR004719; PTSIIC\_glc.  
 DR Pfam; PF00367; PTS EIIB; 1.  
 DR Pfam; PF02378; PTS EIIC; 1.  
 DR TIGRFAMs; TIGR00826; pts-glc; 1.  
 DR TIGRFAMs; TIGR00852; pts-glc; 1.  
 DR PROSITE; PS01035; PTS EIIB\_CYS; 1.  
 KW Phosphotransferase system; Transport; Sugar transport; Transferase; Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 111 EIIB DOMAIN.  
 FT DOMAIN 112 456 EIIC DOMAIN.  
 FT TRANSMEM 112 132 POTENTIAL.  
 FT TRANSMEM 144 164 POTENTIAL.

FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 209 229 POTENTIAL.  
 FT TRANSMEM 247 267 POTENTIAL.  
 FT TRANSMEM 288 308 POTENTIAL.  
 FT TRANSMEM 330 350 POTENTIAL.  
 FT TRANSMEM 360 380 POTENTIAL.  
 FT TRANSMEM 388 408 POTENTIAL.  
 FT TRANSMEM 428 448 POTENTIAL.  
 FT MOD\_RES 25 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 456 AA; 48022 MW; 4AFDF5405CAEPC66 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 456;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 AGWLAGLL 113  
 |||||  
 Db 288 AGWLAGLL 295

RESULT 10  
 PTSSB\_SALTY STANDARD; PRT; 456 AA.  
 AC P08470;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE PTS system, sucrose-specific IIBC component (EIIBC-SCR) (Sucrose-permease IIBC component) (Phosphotransferase enzyme II, BC component) [EC 2.7.1.69] (EIIB-SCR).  
 GN SCRA.  
 OS Salmonella typhimurium.  
 OG Bacterioides pur400.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88216186; PubMed=3285123;  
 RA Ebner R., Lengeler J.W.;  
 RT "DNA sequence of the gene scrA encoding the sucrose transport protein EnzymeII(Scr) of the phosphotransferase system from enteric bacteria: homology of the EnzymeII(Scr) and EnzymeII(Bgl) proteins."; Mol. Microbiol. 2:9-17(1988).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE=94018607; PubMed=8412665;  
 RA Jahreis K., Lengeler J.W.;  
 RT "Molecular analysis of two ScrR repressors and of a ScrR-FurR hybrid repressor for sucrose and D-fructose specific regulons from enteric bacteria."; Mol. Microbiol. 9:195-209(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96188840; PubMed=8628219;  
 RA Titgemeyer F., Jahreis K., Ebner R., Lengeler J.W.;  
 RT "Molecular analysis of the scrA and scrB genes from Klebsiella pneumoniae and plasmid pur400, which encode the sucrose transport protein Enzyme II Scr of the phosphotransferase system and a sucrose-6-phosphate invertase."; Mol. Gen. Genet. 250:197-206(1996).  
 RN [4]  
 RP SEQUENCE OF 1-7 FROM N.A.  
 RC STRAIN=6153-62;  
 RX MEDLINE=91100329; PubMed=1846143;  
 RA Hardesty C., Ferran C., Drenzo J.M.;  
 RT "Plasmid-mediated sucrose metabolism in Escherichia coli: characterization of scrI, the structural gene for a phosphoenolpyruvate-dependent sucrose phosphotransferase system outer membrane porin."; J. Bacteriol. 173:449-456(1991).  
 RL J. Bacteriol. 173:449-456(1991).  
 CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE -TRANSPORT SYSTEM. THE IID DOMAINS CONTAIN THE SUGAR BINDING SITE AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO THE SUGAR. EIIBC-SCR BELONGS TO THE EIIBC DOMAIN TYPE AND LACKS A COVALENTLY BOUND IIA DOMAIN. INSTEAD, EIIB-SCR-MEDIATED PHOSPHORYLATION OF SUCROSE REQUIRES THE ACTIVITY OF ENZYME IIA-GLC, A COMPONENT OF THE MAJOR GLUCOSE TRANSPORT SYSTEM.  
 -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein histidine + sugar phosphate.  
 -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.  
 -!- SIMILARITY: Contains 1 PTS EIIB domain.  
 -!- SIMILARITY: Contains 1 PTS EIIC domain.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 -----  
 CC EMBL; Y00541; CAA68605.1; ALT\_SEQ.  
 DR EMBL; X67750; CAA47973.1; -.  
 DR EMBL; M38416; AAA98418.1; -.  
 DR PIR; S01036; WOEST.  
 DR InterPro; IPR001996; PTS\_EIIB.  
 DR InterPro; IPR003352; PTS\_EIIC.  
 DR InterPro; IPR004719; PTSIIC\_glc.  
 DR Pfam; PF00367; PTS\_EIIB; 1.  
 DR Pfam; PF02378; PTS\_EIIC; 1.  
 DR ProDom; PD001476; PTS\_EIIB; 1.  
 DR TIGRFAMs; TIGR00826; EIIB\_glc; 1.  
 DR TIGRFAMs; TIGR00852; pts-glc; 1.  
 DR PROSITE; PS01035; PTS\_EIIB\_CYS; 1.  
 KW Phosphotransferase system; Transferrase; Transport; Sugar transport; Plasmid; Transmembrane; Inner membrane; Phosphorylation.  
 FT DOMAIN 1 111 EIIB DOMAIN.  
 FT DOMAIN 112 456 EIIC DOMAIN.  
 FT TRANSMEM 144 164 POTENTIAL.  
 FT TRANSMEM 181 201 POTENTIAL.  
 FT TRANSMEM 213 233 POTENTIAL.  
 FT TRANSMEM 247 267 POTENTIAL.  
 FT TRANSMEM 288 308 POTENTIAL.  
 FT TRANSMEM 330 350 POTENTIAL.  
 FT TRANSMEM 360 380 POTENTIAL.  
 FT TRANSMEM 388 408 POTENTIAL.  
 FT TRANSMEM 428 448 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 25 26 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 309 309 PHOSPHORYLATION (BY SIMILARITY).  
 SQ SEQUENCE 456 AA; 47864 MW; 98A6F1620AE50885 CRC64;

Query Match 1.4%; Score 8; DB 1; Length 456;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 AGWLAGLL 113  
 |||||  
 Db 288 AGWLAGLL 295

RESULT 11  
 NLDDL RAT STANDARD; PRT; 745 AA.  
 AC O54697;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE N-acetylated-alpha-linked acidic dipeptidase like protein (EC 3.4.17.21) (NAALADase L) (Ileal dipeptidylpeptidase) (100 kDa ileum brush border membrane protein) (I100).  
 DE





```
FT DOMAIN 935 948 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 949 968 POTENTIAL.
FT DOMAIN 969 998 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 999 1020 POTENTIAL..
FT DOMAIN 1021 1027 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1028 1050 POTENTIAL.
FT DOMAIN 1051 1056 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1057 1077 POTENTIAL.
FT DOMAIN 1078 1090 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1091 1115 POTENTIAL.
FT DOMAIN 1116 1158 CYTOPLASMIC (POTENTIAL).
FT MOD RES 457 457 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 859 859 MAGNESIUM (BY SIMILARITY).
FT METAL 863 863 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1158 AA; 130328 MW; 5CC042B40C8C974D CRC64;

Query Match 1.4%; Score 8; DB 1; Length 1158;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 ILLSLFLI 103
DB 331 ILLSLFLI 338

RESULT 13
MAX7_BOMMX STANDARD; PRT; 20 AA.
AC P83086;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Maximin 7.
OS Bombina maxima (Giant fire-bellied toad) (Chinese red belly toad).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Archeobatrachia; Bombinatoridae; Bombina.
OX NCBI_TaxID=161274;
RN [1]
RP SEQUENCE, FUNCTION, SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RC TISSUE=Skin secretion;
RA Chen T.B., McClean S., Orr D.F., Bjourson A.J., Rao P.F., Shaw C.;
RT "Isolation and structural characterisation of antimicrobial peptides
RT from the venom of the Chinese large-webbed bell toad (Bombina
RT maxima).";
RL Submitted (JUL-2001) to the SWISS-PROT data bank.
CC -!- FUNCTION: Has antimicrobial activity.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: Skin.
CC -!- TISSUE SPECIFICITY: Skin.
KW Amphibian defense peptide; Antibiotic.
SQ SEQUENCE 20 AA; 1947 MW; EC7541AA1BF3BD49 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 9.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 60 GPVLGLV 56
DB 3 GPVLGLV 9

RESULT 14
EX7S_PARZE STANDARD; PRT; 78 AA.
AC Q8LIH9;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable exodeoxyribonuclease VII small subunit (EC 3.1.11.6)
DE (Exonuclease VII small subunit).
GN XSEB.
OS Paracoccus zeaxanthinifaciens.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
```

```
OC Rhodobacteraceae; Paracoccus.
OX NCBI_TaxID=187400;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI14;
RA Huemelin M.;
RT "Genetics of isoprenoid biosynthesis in Paracoccus
RT zeaxanthinifaciens.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Bidirectionally degrades single-stranded DNA into large
CC acid-insoluble oligonucleotides, which are then degraded further
CC into small acid-soluble oligonucleotides (By similarity).
CC -!- CATALYTIC ACTIVITY: Exonucleolytic cleavage in either 5'- to 3'-
CC or 3'- to 5'-direction to yield nucleoside 5'-phosphates.
CC -!- SUBUNIT: Heterooligomer composed of large and small subunits (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE XSEB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ431697; CAD24424.1; -.
DR HAMAP; MF_00337; -; 1.
DR InterPro; IPR003761; Exonuc VII_S.
DR Pfam; PF02609; Exonuc VII_S; 1.
DR TIGRFAMs; TIGR01280; xseB; 1.
KW Hydrolase; Nuclease; Exonuclease.
SQ SEQUENCE 78 AA; 8492 MW; 1D3963A65D5C53D8 CRC64;

Query Match 1.3%; Score 7; DB 1; Length 78;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 TEPAEGL 232
DB 72 TEPAEGL 78

RESULT 15
CYC6_SPIWA STANDARD; PRT; 89 AA.
AC P00118;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cytochrome c6 (Soluble cytochrome f) (Cytochrome c553) (Cytochrome c-
DE 553).
GN PETJ.
OS Spirulina maxima (Arthrospira maxima).
OC Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.
OX NCBI_TaxID=1155;
RN [1]
RP SEQUENCE.
RX MEDLINE=75100362; PubMed=803642;
RA Ambler R.P., Bartsch R.G.;
RT "Amino acid sequence similarity between cytochrome f from a
RT blue-green bacterium and algal chloroplasts.";
RL Nature 253:285-288 (1975).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS).
RX MEDLINE=21371781; PubMed=11478889;
RA Sawaya M.R., Krogmann D.W., Serag A., Ho K.K., Yeates T.O.,
RA Kerfeld C.A.;
RT "Structures of cytochrome c-549 and cytochrome c6 from the
RT cyanobacterium Arthrospira maxima.";
RL Biochemistry 40:9215-9225 (2001).
RN [3]
```



---

**THIS PAGE BLANK (USPTO)**

---

**THIS PAGE BLANK (USPTO)**